

UERJ

Dissertação de Mestrado em Engenharia de Computação

CLASSIFICAÇÃO DE IMAGENS UTILIZANDO CARACTERÍSTICAS DE TEXTURA

Autor: Marcelo Musci Zaib Antonio

Orientadora: Maria Luiza Fernandes Velloso

Co-orientador: Flávio Joaquim de Souza

Programa de Pós-Graduação em Engenharia de Computação

Área de Concentração: Geomática



Faculdade de Engenharia

CLASSIFICAÇÃO DE IMAGENS UTILIZANDO CARACTERÍSTICAS DE TEXTURA

Marcelo Musci Zaib Antonio

Dissertação submetida ao corpo docente da Faculdade de Engenharia da Universidade do Estado do Rio de Janeiro – UERJ, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Computação – Área de Concentração Geomática.

Orientadora: Maria Luiza Fernandes Velloso

Co-orientador: Flávio Joaquim de Souza

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO

Rio de Janeiro
Março – 2008

MUSCI, MARCELO ZAIB ANTONIO.

Classificação de Imagens Utilizando Características de Textura [Rio de Janeiro] 2008.

viii, 111 p. 29,7 cm (FEN/UERJ, M.Sc., Programa de Pós-Graduação em Engenharia de Computação – Área de Concentração Geomática, 2008)

Dissertação – Universidade do Estado do Rio de Janeiro – UERJ

1 – Processamento Digital de Imagens

2 – Análise de Texturas

3 – Inteligência Computacional

I. FEN/UERJ II. Título (série)

FOLHA DE JULGAMENTO

Título: Classificação de Imagens Utilizando Características de Textura.

Candidato: Marcelo Musci Zaib Antonio

Programa: Programa de Pós-Graduação em Engenharia de Computação – Área de Concentração Geomática

Data da defesa: 19 de março de 2008.

Aprovada por:

Orientadora: Maria Luiza Fernandes Velloso, DSc., UERJ.

Co-orientador: Flávio Joaquim de Souza, DSc., UERJ.

Orlando Bernardo Filho, DSc., UERJ.

Fabiano Saldanha Gomes de Oliveira, DSc., LNCC.

AGRADECIMENTOS

Acima de tudo, agradeço a Deus pela dádiva da vida e por, em muitos momentos aflitivos, proporcionar-me a sua paz e a serenidade para enfrentar os obstáculos que me atravessavam e superar os desafios.

À minha orientadora, Maria Luiza Fernandes Velloso, quem admiro não só pela inteligência e intelectualidade, mas, especialmente, pelos valorosos conselhos acadêmicos ao longo da graduação e mestrado.

A todas as pessoas que, direta ou indiretamente, contribuíram para a realização desse trabalho monográfico.

Resumo da Dissertação apresentada à FEN/UERJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.).

CLASSIFICAÇÃO DE IMAGENS UTILIZANDO CARACTERÍSTICAS DE TEXTURA

Marcelo Musci Zaib Antonio

Março/2008

Orientadora: Maria Luiza Fernandes Velloso, DSc., UERJ.

Co-orientador: Flávio Joaquim de Souza, DSc., UERJ.

Programa de Pós-Graduação em Engenharia da Computação – Área de Concentração Geomática.

Essa dissertação tem o objetivo de verificar a contribuição da análise textural à classificação de imagens multiespectrais, com o possível uso na discriminação e mapeamento de classes de cobertura da terra. Nesse contexto, o trabalho proposto efetua uma comparação entre dois diferentes métodos de extração de características de textura, a saber, matrizes de co-ocorrência (GLCM, *Gray Level Co-occurrence Matrix*) e distribuição estatística de formas geométricas (SGF, *Statistical Geometric Features*), utilizando resultados experimentais, em imagens sintéticas e de sensoriamento, para demonstrar o desempenho dos métodos citados. No processo de classificação supervisionada de imagens, com a utilização de redes neurais, foi avaliado o uso da rede neural probabilística (PNN, *Probabilistic Neural Network*) apresentando desempenho superior às redes perceptrons multicamadas (MLP, *Multi Layer Perceptron*).

Palavras-chave: Análise de Textura, Processamento Digital de Imagem, Redes Neurais

Abstract of the dissertation presented to FEN/UERJ as part of the necessary requirements to obtain the degree of Master in Science.

IMAGE CLASSIFICATION USING TEXTURE FEATURES

Marcelo Musci Zaib Antonio

Março/2008

Advisors: Maria Luiza Fernandes Velloso, DSc., UERJ

Flávio Joaquim de Souza, DSc., UERJ.

Post Graduation Program in Computer Engineering – Field: Geomatics.

This work investigates the contribution of the textural analysis to the classification of multispectral images, and its use in land cover mapping. Two different methods of extraction of texture characteristics were compared: GLCM -Gray Level Co-ocurrence Matrix and SGF - Statistical Geometric Features. In order to assess and compare the effectiveness of the used techniques, experiments were carried on two different data sets. One was a combination of Mit's VisTex texture image patterns, widely used for texture analysis comparisons. The other was a data set of real remotely sensed patterns. Texture classification involves deciding to which texture category an observed image belongs. Once the texture features are extracted, one then uses pattern classification techniques in order to classify. Two techniques of supervised classification were used and compared, the MLP- Multi-Layer Perceptron neural network and the PNN - Probabilistic Neural Network, which presented superior performance.

Keywords: Texture Analysis, Image Digital Processing, Neural Network..

SUMÁRIO

Agradecimentos	i
Resumo	ii
Abstract	iii
Lista de abreviaturas e siglas	vii
Lista de símbolos	viii
1 Introdução	1
1.1 Motivação	1
1.2 Objetivo	3
1.3 Descrição do Trabalho	4
2 Reconhecimento de Padrões	5
2.1 Introdução	5
2.2 Pré-processamento de Imagem	8
2.3 Descrição e Medição de Características	11
2.3.1 Características Multi-espectrais	12
2.3.1.1 Modelo de Cores	12
2.3.2 Características de Textura	13
2.3.2.1 Informações de Cor em Textura	14
2.3.3 Medição das Características	15
2.3.3.1 Aplicação de Heurística para Melhorar o Desempenho de Classificação	17
2.4 Extração de Características	18
2.4.1 Análise de Componentes Principais (PCA – <i>Principal Components Analysis</i>)	20
2.5 Pré-processamento de características	21
2.6 Seleção de Características	22
2.6.1 Critérios para Seleção de Características	22
2.6.2 Algoritmos de Busca para Seleção de Características	24

2.7 Classificação	26
2.7.1 Classificação não-supervisionada	27
2.7.2 Classificação supervisionada	27
2.7.3 Estimativa de Erro	31
3 Quantização de Textura	32
3.1 Introdução	32
3.2 Matriz de Co-ocorrência	32
3.2.1 Matriz de Co-ocorrência de Níveis de Cinza – GLCM	33
3.3 Distribuição Estatística de Formas Geométricas – SGF	44
3.3.1 Algoritmo da Distribuição Estatística de Formas Geométricas	45
4 Redes Neurais para Classificação	49
4.1 Introdução	49
4.2 Rede Neural Multicamadas	51
4.3 Rede Neural de Função de Base Radial	53
4.3.1 Rede Neural Probabilística – PNN	55
4.4 Metodologia de Construção das Redes Neurais Utilizadas	56
4.4.1 Coleta de Dados e Separação em Conjuntos	56
4.4.2 Configuração da rede	57
4.4.3 Treinamento	59
4.4.4 Teste	58
5 Testes e Experimentos	59
5.1 Introdução	59
5.2 Escolha do Classificador	61
5.3 Avaliação e Testes com GLCM	65
5.3.1 Reduzindo o Espaço de Entrada de GLCM.....	66
5.4 Avaliação e Testes com SGF	71
5.4.1 Reduzindo o Espaço de Entrada de SGF	73
5.5 Comparação dos Métodos GLCM e SGF	75
5.6 Uso de Cores	76
6 Conclusões.....	80
Referências	82
A Fundamentação Estatística.....	87
A.1 Análise de Componentes Principais (PCA - <i>Principal Components Analysis</i>)	88

A.2 Distancia entre Distribuições	90
A.3 Coeficiente de Kappa	93
B Classificadores Usados.....	96
B.1 Rede Neural Multicamadas (MLP – <i>Multilayer Perceptron</i>)	97
B.2 Rede Neural Probabilística (PNN – <i>Probabilistic Neural Network</i>)	102
C Conectividade e Medidas de Irregularidades	115
C.1 Conectividade	106
C.2 Medidas de Irregularidades	107
D Biblioteca VisTex	110

LISTA DE ABREVIATURAS E SIGLAS

GLCM - Matriz de Co-ocorrência do Nível de Cinza (*Gray Level Co-occurrence Matrix*)

HSV – Modelo de cores H - matiz (*Hue*), S - saturação (*Saturation*), V - intensidade (*Value*)

MLP – Multicamadas perceptron (Multi Layer Perceptron)

PCA - Análise de Componentes Principais (*Principal Components Analysis*)

pdf – função de densidade de probabilidade (*probability density function*)

PNN – Rede Neural Probabilística (*Probabilistic Neural Network*)

RBF – Rede Neural de Base Radial (*Radial Basis Function*)

RNA – Rede Neural Artificial

RP – Reconhecimento de Padrões

SGF - Distribuição Estatística de Formas Geométricas (*Statistical Geometric Features*)

LISTA DE SÍMBOLOS

A, B	conjunto de pontos no espaço Euclidiano,
c, v, s, g	índices,
i, j, d	índices de matrizes,
k, l, m, n	pontos no \mathbb{Z}^2 ,
(x, y)	pontos no espaço Euclidiano bidimensional,
ρ	matriz de covariância,
σ	desvio padrão,
\mathbb{R}	números reais,
\mathbb{Z}	conjunto dos inteiros
\mathbb{R}^k	espaço Euclidiano k -dimensional,
\mathbb{Z}^k	espaço discreto k -dimensional,
D, G, J	subconjuntos do espaço Euclidiano; domínio de uma função,
f	função,
I	imagem,
$\#$	número de,
$ A $	cardinalidade do conjunto A ,
$\ \cdot\ $	norma Euclidiana,
$[a, b]$	intervalo fechado de uma variável x , significa $a \leq x \leq b$
$\forall x$	para todo x
$\exists x$	existe um x tal que,
\Rightarrow	implica,
$A \subset B$	A está incluído em B ,
$A \supset B$	A contém B ,
$a \in A$	ponto a pertence ao conjunto A ,
$a \notin A$	ponto a não pertence ao conjunto A .

Capítulo 1

INTRODUÇÃO

1.1 Motivação

O levantamento do uso da terra numa dada região tornou-se um aspecto fundamental para a compreensão dos padrões de organização do espaço. Entretanto, as medidas para o planejamento do uso da terra têm sido, até recentemente, baseadas em informações fragmentadas sobre os efeitos do uso do solo no ambiente. Isto ocorria porque não existia registro seguro sobre as condições de uso da terra, não se podendo avaliar as alterações que são provocadas pelo homem. A utilização de dados atualizados de uso e revestimento da terra é muito ampla, podendo-se citar, por exemplo: inventário de recursos hídricos, controle de inundações, identificação de áreas com processos erosivos avançados, ocupação urbana, avaliação de impactos ambientais, formulação de políticas econômicas, etc.

Por outro lado, para a caracterização do uso atual das terras existem vários métodos tradicionais de levantamento de campo, como aqueles auxiliados por fotografias aéreas ou pela cartografia temática, entre outros. Estes métodos são caros e de difícil execução, para resultados que, muitas vezes, ao final do trabalho já estão desatualizados. Sendo assim, o uso de métodos tradicionais de levantamento do uso das terras, em razão do custo elevado dos instrumentos e a dificuldade de disponibilizar as informações de maneira rápida aos tomadores de decisão, torna proibitivo seu emprego de forma sistemática e repetitiva sobre grandes extensões de território. Desta forma, existe a necessidade de se utilizar métodos que possibilitem o levantamento do uso das terras de maneira eficiente, rápida e que tenham relativamente baixo custo.

Neste contexto, o sensoriamento remoto constitui-se numa técnica de grande utilidade, pois permite em curto espaço de tempo a obtenção de uma grande quantidade de informações a respeito de registros de uso da terra. O levantamento do uso da terra numa dada região tornou-se um aspecto de interesse fundamental para a compreensão dos padrões de organização do espaço, (Santos et al.,1981).

O advento dos satélites de sensoriamento remoto tem favorecido, nos últimos anos, a realização de levantamentos à distância de variações físicas e químicas da superfície terrestre em áreas extensas e inóspitas, favorecendo, principalmente, os mapeamentos e monitoramentos sazonais da superfície da Terra.

Para extrair informações de imagens de sensores remotos o intérprete considera, na interpretação visual, características como o contexto, formas, bordas, variações tonais ou cores, e textura.

Na interpretação digital, a maioria das técnicas-padrão de classificação disponíveis apresentam algoritmos de classificação baseados somente no uso da informação espectral, ou seja, na intensidade de cinza ou características tonais, deixando de lado as outras características consideradas na interpretação visual. Mas, pesquisas vêm sendo efetuadas para incluir estas outras características para a extração automática de informações de imagens.

A textura pode ser pensada em termos de suas características como um sub-padrão ou primitiva, e por uma regra de localização que define sua distribuição espacial; ou também se poderia pensar a textura como um fenômeno organizado composto de padrões que obedecem alguma regra estatística, como citado por Haralick & Shapiro (1992) e Visa (1990).

No processamento digital de imagens de sensoriamento remoto, as feições texturais contêm informações sobre a distribuição espacial das variações tonais dentro de uma banda, e as feições espectrais descrevem a variação tonal das várias bandas de uma imagem. Enquanto a classificação multi-espectral¹ é um processo de agrupamento de *pixels* de valores de intensidade espectral igual ou muito próximos, na análise textural, ao contrário, a diferença das feições tonais é que irão caracterizá-la, portanto ela é expressa pela diferença entre as intensidades dos vizinhos, e será sempre fundamentada em uma operação de área ou vizinhança, Crosta (1992).

Textura é uma propriedade inata das imagens e contém informações importantes a respeito da disposição estrutural das superfícies. O uso de textura, além das características espectrais, para classificação de imagens pode trazer resultados mais precisos, dependendo apenas da resolução espacial do sensor e o quão homogênea é a área a ser classificada.

¹ Imagens de várias bandas de comprimento de ondas.

1.2 Objetivo

O presente trabalho dessa dissertação foi o de pesquisar e avaliar metodologias de classificação de imagens com a utilização de análise textural e o uso de redes neurais, propondo uma metodologia para uso em sistemas de classificação de imagens oriundas de sensoriamento remoto.

Entre as aplicações que vêm utilizando a análise textural de dados espectrais, Peddle & Franklin (1998) citam: análise da Terra; mapeamento florestal; monitoramento ambiental e estudos ecológicos.

Várias técnicas foram propostas para a utilização na extração de características texturais em imagens; para Schowengerdt (1983), não existe uma caracterização definitiva de textura, mas, todas as definições numéricas são descritas para rugosidade e contraste dos detalhes da imagem. Ele define textura como uma propriedade local de uma imagem, i.e., uma medida da intensidade de cinza entre os *pixels* e sua vizinhança. Dessa forma podemos extrair características de texturas em imagens de diversas formas. Ao longo do trabalho serão feitas comparações entre algumas técnicas mais usadas.

Diversos métodos de classificação supervisionada podem ser utilizados, sendo os mais usados aqueles baseados no uso de redes neurais artificiais (RNAs) - discutidas nos Capítulo 2 e Capítulo 4.

Existem vários modelos de redes neurais artificiais utilizadas nas aplicações de classificação de imagens oriundas de sensoriamento remoto. Foram empregados no corrente trabalho dois modelos de RNAs - Rede Perceptron Multicamada (MLP – *Multi Layer Perceptron*) e Rede Neural Probabilística (PNN – *Probabilistic Neural Network*).

Diversos trabalhos foram efetuados utilizando esses tipos de rede, a citar os trabalhos de Raghu, P.P. & Yegnanarayana, B. (1998) e Ramakrishnan, A.G. et al. (2002).

Nesse contexto, o trabalho proposto possui o objetivo de verificar a contribuição da análise textural à classificação espectral de imagens, com o possível uso na discriminação e mapeamento de classes de cobertura da terra, fazendo uma comparação entre diferentes métodos de extração de características de textura, além de comparações entre os métodos de classificação supervisionada, com a utilização de redes neurais.

Como a interpretação das imagens de satélite é uma tarefa que exige que o fotointérprete tenha, além de bastante experiência na classificação de imagens, um conhecimento de campo bastante amplo. É visando facilitar o trabalho deste profissional que essa dissertação efetua uma análise de técnicas distintas de forma a melhorar o desempenho no uso de texturas e RNAs em classificação de imagens.

1.3 Descrição do Trabalho

Esta dissertação desdobra-se em seis capítulos. Neste capítulo 1 são apresentadas as motivações e objetivos. Nos capítulos 2 e 3 é, de forma objetiva, apresentada a fundamentação teórica das duas principais áreas do conhecimento envolvidas no trabalho: no capítulo 2 a introdução teórica sobre reconhecimento de padrões e no capítulo 3 a introdução teórica sobre texturas. No capítulo 4 são apresentados os tipos de redes neurais usados neste trabalho. No capítulo 5 são apresentados os testes e experimentos realizados, com os resultados obtidos. Ao final, no capítulo 6, são apresentadas as conclusões sobre o trabalho.

Capítulo 2

RECONHECIMENTO DE PADRÕES

2.1 Introdução

Reconhecimento de padrões (RP) é a aplicação de técnicas matemáticas e estatísticas para a identificação e classificação de objetos de diferentes classes. Em geral, RP é uma ciência relativa à descrição e reconhecimento de objetos usando algum tipo de medida.

Conforme o tipo de objetos a classificar (ou descrever) o projeto de RP usa algum (ou alguns) dos seguintes principais métodos ou abordagens: estatística, estrutural, sintática e mais recentemente redes neurais. Técnicas de RP são geralmente desenvolvidas em dois estágios, a saber: treinamento, onde sistemas RP são inicialmente projetados e teste, onde o desempenho do sistema é avaliado. Técnicas de RP podem ser vistas em Theodoridis & Koutroumbas (2003) e Sing-Tze-Bow (2002).

Dois elementos essenciais dentro de reconhecimento de padrões devem ser definidos, (Pandya, 1995):

- **Padrão:** É um conjunto de características² que definem um objeto ou um grupo de objetos. É essencialmente um arranjo ou uma ordenação, em que existe alguma organização estruturada. Um padrão pode ser referenciado como uma quantidade ou descrição estrutural de um objeto ou algum outro item de interesse. Um padrão pode ser tão básico quanto um conjunto de medidas ou observações, geralmente sendo representado na forma de vetor ou matriz.
- **Características:** Podem ser entendidas como qualquer medição útil extraída no processo de identificação do padrão. As características podem ser simbólicas, numéricas ou ambas. Podem ser variáveis contínuas ou discretas.

Segundo Souza (1999), problemas do mundo real podem ser representados por \mathbf{R} características. Esta é a dimensionalidade do espaço dos padrões. Como \mathbf{R} é, geralmente,

² Alguns autores preferem o termo “*descritores*”.

muito grande, é desejável reduzi-lo (*redução da dimensionalidade*) de forma que os dados resultantes ainda mantenham o poder discriminatório dos padrões que estão inerentes aos dados. O espaço das características é postulado de dimensão \mathbf{N} ($\mathbf{N} < \mathbf{R}$) no qual as regras de classificação podem ser executadas em tempo razoável. O espaço de classificação, então, simplesmente é o espaço de decisão no qual \mathbf{K} classes podem ser selecionadas e, portanto, de dimensão \mathbf{K} . Conceitualmente, o problema de RP pode, então, ser descrito como uma transferência do espaço de padrões \mathbf{P} (dimensão \mathbf{R}), para o espaço de características \mathbf{F} (dimensão \mathbf{N}) e finalmente para o espaço de classificação \mathbf{C} (dimensão \mathbf{K}):

$$\mathbf{P} \rightarrow \mathbf{F} \rightarrow \mathbf{C} \quad (1)$$

Estas transformações são altamente não lineares e não inversíveis, mas o único objetivo é manter o poder discriminatório dos padrões.

Neste processo encontram-se duas transformações:

- 1º) redução de dimensionalidade;
- 2º) classificação dos dados em \mathbf{K} classes.

Entende-se por classes, grupos de objetos cujas propriedades intrínsecas diferem entre si. Podemos caracterizar classes tomando medidas qualitativas ou quantitativas das propriedades dos objetos, ou seja, suas características. Podem ser utilizados diferentes métodos para descrição, assim como, de mensurar as características dos objetos. A princípio não sabemos quais características dos objetos melhor poderá ser utilizada para uma boa discriminação entre as classes envolvidas, portanto é comum se extrair uma grande quantidade de características inerentes aos objetos envolvidos durante a fase de treinamento do sistema de classificação. Esse conjunto formado pelas características dos objetos é então reduzido por dois processos conhecidos como *extração de características (feature extraction)* e *seleção de características (feature selection)*. Esses processos envolvem parâmetros de medidas estatísticas (como média e variância) originadas das características de distribuição das classes envolvidas. As características que possibilitam uma melhor separação entre as classes envolvidas são ditas possuírem *alta discriminação*.

O objetivo de qualquer sistema de reconhecimento de padrões é alocar um determinado objeto, cuja classe seja desconhecida, em uma dentre varias outras classes. Classificar uma área montanhosa de uma planície é um exemplo. A alocação de objetos em classes, baseado nas suas características, é chamado de *classificação*. A construção de classificadores é

baseada em *análise discriminatória* – técnica de separação e classificação de dois ou mais grupos de dados. O processo de análise discriminatória é formado por duas etapas: A primeira consiste em discriminar entre grupos de dados multi-variados originados de fontes conhecidas. Esses grupos de dados são os grupos de controle, conhecidos como *conjunto de treinamento*. A segunda etapa é a classificação dos dados de origem desconhecida entre um dos grupos envolvidos, aplicando uma função de discriminação.

O modelo de sistema proposto para classificação de imagens, utilizando características de textura, é ilustrado na Figura 1, que é composto pelos seguintes módulos principais:

- Aquisição de Imagem;
- Pré-processamento, que corrige eventuais falhas ou distorções na imagem, além de efetuar uma quantização³ dos níveis de cinza da imagem;
- Descrição e Medição de Características, toma medidas da imagem para posterior classificação;
- Extração de Características, cria um vetor de características (*feature vector*) reduzido, com dados medidos da imagem;
- Pré-processamento de Características, transformação dos dados de forma a serem melhor utilizados;
- Seleção de Características, elimina informações redundantes;
- Classificação, toma por base os padrões extraídos da imagem e efetua a classificação da mesma.

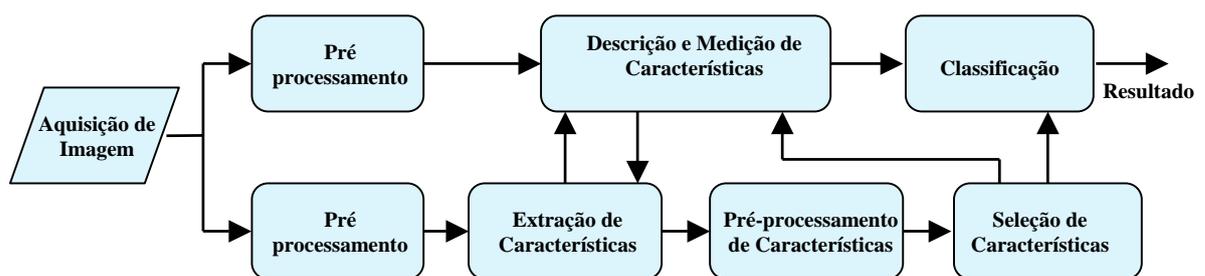


Figura 1 – Componentes fundamentais de um sistema de reconhecimento de padrões utilizado para o processamento de imagens.

³ Processo de atribuição de valores discretos para um sinal cuja amplitude varia entre infinitos valores.

2.2 Pré-processamento de Imagem

O pré-processamento de imagens digitais é uma necessidade em muitos sistemas de análise, sendo às vezes referidas como restauração de imagem e retificação. A aplicação do pré-processamento torna-se necessária por diversas razões, incluindo:

- restaurar imagens que tenham sido corrompidas de algum modo;
- aumentar atributos de imagens que forem de interesse particular;
- normalização ou quantização das imagens de forma a melhorar a sua qualidade visual.

Imagens corrompidas por “impulsos de ruídos”⁴ podem ser melhoradas com a utilização de filtros medianos. Um filtro mediano repõe o nível de cinza de um *pixel* pelo valor médio dos níveis de cinza dos *pixels* vizinhos. Remoção de ruído é um exemplo de restauração de imagem, podendo ser observado na Figura 2.

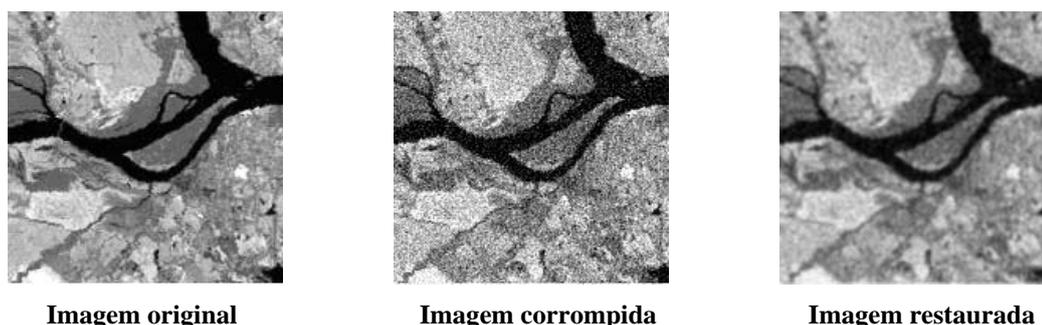


Figura 2 – Processo de restauração de uma imagem corrompida por impulso de ruído.

A imagem restaurada é o resultado da aplicação de um filtro mediano na imagem corrompida.

Fonte: *Natural Resources Canadá*.

Um segundo tipo de pré-processamento, chamado realce de imagem, envolve a aplicação de operadores de forma a aumentar ou destacar atributos da imagem. Realces são utilizados para facilitar a interpretação visual e a compreensão da imagem. Um exemplo de realce é a manipulação do contraste de uma imagem, utilizando a correção do histograma da imagem, de forma a aumentar a faixa dinâmica dos níveis de cinza da imagem.

⁴ Impulsos de ruído - processo estocástico que resulta em uma distribuição aleatória de pixels na imagem com alta intensidade.

Tipicamente, as imagens possuem valores de intensidade de *pixels* correspondente a um intervalo máximo que vai de 0 até 255. Devido à má iluminação, defeitos do sensor ou mesmo às características da cena, o intervalo de valores de intensidade ocupados pelos *pixels* presentes em uma imagem, pode ser muito menor que esse intervalo máximo. Diz-se então que a imagem possui baixo contraste, o que torna difícil a sua visualização ou interpretação por um intérprete humano ou um sistema de processamento digital.

O contraste expande a pequena faixa para os valores de 0 à 255. Isto acentua o contraste na imagem, proporcionando um aumento das áreas luminosas e ao mesmo tempo fazendo as áreas escuras aparecendo mais escuras, tornando assim a interpretação visual muito mais fácil.

A Figura 3 ilustra o aumento do contraste em uma imagem. No exemplo, o valor de mínimo (ocupado por dados reais) no histograma é 84 e o valor de máximo é 153. Esses 70 níveis ocupam menos que um terço da faixa total de 256 níveis disponíveis.

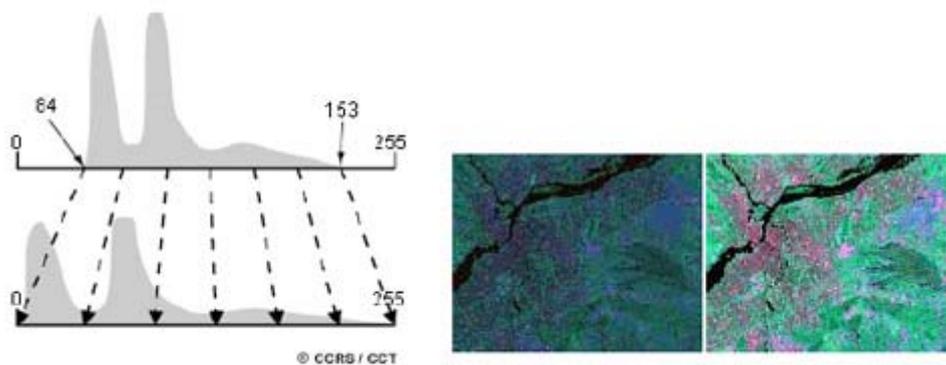


Figura 3 - Operação de manipulação de contraste, aumento do contraste em uma imagem antes (esquerda) e depois (direita) do contraste.

Fonte: *Natural Resources Canadá*.

Normalização de imagens é outro exemplo de pré-processamento, que envolve a utilização de algoritmos para quantização e equalização de histograma. Operações de quantização permitem a redução de um grande número de intensidade de *pixels* variados de uma imagem para um número pequeno e fixo. As dificuldades computacionais no processamento imposto por esses algoritmos estão diretamente relacionadas ao número de intensidades presentes na imagem original N_g . Em imagens geralmente são utilizados 8 *bits* ($N_g = 256$) para representar o nível de intensidade de cinza. Pela re-quantização das imagens com $N_g = 256$ para $N_g = 16$ níveis de cinza, podemos reduzir substancialmente os recursos computacionais necessários, sem depreciar o processo de análise das imagens.

Existem diversas formas de quantização que podem ser utilizadas em imagens, a saber: quantização uniforme, quantização gaussiana e quantização de igual probabilidade, (Haralick, 1979).

A quantização uniforme é o modo mais simples, na qual os níveis de cinza são quantizados com tolerância uniforme entre os mesmos, sem qualquer consideração a respeito da distribuição dos níveis de cinza da imagem em análise. A quantização gaussiana assume que a distribuição dos níveis de cinza da imagem possuem distribuição normal. Cada nível de cinza possui a mesma área sob a curva. A quantização de igual probabilidade é muito utilizada e assume que cada nível de cinza possui uma mesma probabilidade de ocorrer na imagem.

O esquema de quantização gaussiana assume que existe normalidade na distribuição dos níveis de cinza da imagem, o que não ocorre nos tipos de imagem que serão analisadas nesse trabalho. A quantização de igual probabilidade, normaliza diferentes amostras da imagem, de forma que partes escuras e claras assumem a mesma textura e possuem a mesma GLCM, o que não é desejado em algumas situações de análise de texturas envolvendo regiões com água, (Barber e LeDrew, 1991). Em virtude dessas considerações, nas experiências desse trabalho será utilizado o esquema de quantização uniforme.

Para o processo de quantização foi utilizada a fórmula dada em (2), onde o par (x,y) é um *pixel*, $I'(x,y)$ representa o valor re-quantizado de intensidade $I(x,y)$, e a função $\text{floor}(\cdot)$ reduz o valor real dessa quantidade para o maior inteiro abaixo do argumento.

$$\forall(x, y) = \text{floor} \left\{ \frac{I(x, y) - I_{\min}}{I_{\max} - I_{\min}} \cdot N_g \right\} + 1, \quad (2)$$

A adição de 1 na equação (2) resulta em uma faixa de intensidade re-quantizada de $I'(x,y) \in 1, \dots, N_g$, (Walker, 1997).

Na Figura 4 pode ser visto em (a), o histograma de uma imagem com todos os níveis de cinza originais, enquanto em (b) é mostrado o histograma re-quantizado para 16 níveis, pode ser observado que a quantização uniforme não altera o histograma da imagem original.

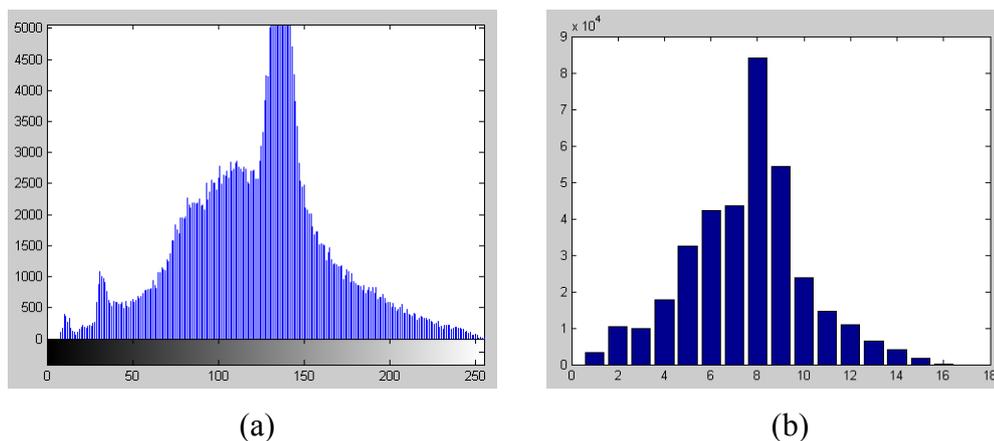


Figura 4 – Histogramas de uma imagem, em (a) histograma em todos os níveis de cinza original da imagem e em (b) o histograma re-quantizado para 16 níveis de cinza.

2.3 Descrição e Medição de Características

Características de uma imagem são propriedades ou atributos dessa imagem. Algumas características são ditas naturais no sentido de que são definidas pela aparência visual da imagem a qual pertencem, enquanto outras características ditam artificiais resultam de manipulações específicas da imagem. Características naturais incluem a luminância⁵ dos *pixels* da imagem e a escala de cinza de uma região de textura. Histogramas de amplitude e frequências espectrais são exemplos de características artificiais.

As características são de suma importância para o completo isolamento de regiões na imagem que possuem propriedades comuns (*segmentação de imagens*) e a subsequente identificação dessas regiões (*classificação de imagens*).

Os tipos de características mais utilizadas em classificação de imagens recaem geralmente nas seguintes categorias:

- Características morfológicas – propriedades relacionadas com a forma e o tamanho dos objetos da imagem.
- Características multi-espectrais – propriedades de cores e frequência.
- Características de textura – relacionamento espacial entre as intensidades da imagem.

⁵ luminância – medida de luminosidade total da imagem.

Foram utilizados, nesse trabalho, técnicas de descrição de características para classificação de imagens baseadas em *características multi-espectrais* e *características de textura*.

2.3.1 Características Multi-espectrais

Podem ser definidas como as informações de cor dos *pixels* de uma imagem, sendo cada cor composta de informação na forma de comprimentos de onda (frequência de cor). Podemos compor a informação de cor usando três cores primárias: vermelho, verde e azul ou em inglês *red* (R), *green* (G) e *blue* (B), respectivamente.

As características normalmente usadas para distinguir uma cor da outra são *brilho*, *matiz* e *saturação*. O brilho incorpora a noção cromática da intensidade. Matiz é um atributo associado com o comprimento de onda dominante em uma mistura de ondas de luz. A saturação refere-se a pureza relativa ou quantidade de luz branca misturada com um matiz. O matiz e a saturação, quando tomados juntamente, são chamados *cromaticidade* ou informação de *croma*, e portanto, uma cor pode ser caracterizada pelo seu brilho e cromaticidade, (González & Woods, 2002).

2.3.1.1 Modelo de Cores

De acordo com Gonzalez & Woods (2002), o propósito de um modelo de cores é facilitar a especificação das cores em alguma forma padrão e de aceitação geral. Essencialmente, um modelo de cor é uma especificação de um sistema de coordenadas tridimensionais e um subespaço dentro desse sistema onde cada cor é representada por um único ponto.

Os modelos de cores mais frequentemente usados para processamento de imagens são o RGB, o YIQ, o HSV e o LUV.

Diversos trabalhos utilizando esses modelos foram utilizados, sendo no caso dessa dissertação foi escolhido o modelo HSV de cores. Nesse modelo as informações de cromaticidade são representadas pelos componentes H (*Hue* – matiz) e S (*Saturation* – saturação) e a informação de intensidade pela componente V (*Value* – valor). Se fizermos os valores de H e S iguais a zero, então a componente V irá representar uma imagem em tons de cinza (*grayscale*).

2.3.2 Características de Textura

Nos últimos anos vem crescendo a utilização do uso de características de textura na classificação de imagens, segundo Schowengerdt (1983), a textura pode ser medida por quantidades estatísticas simples, como a variância do nível de cinza local, Barber & LeDrew (1991) comentam que existem várias abordagens para medidas de textura, incluindo: estatística, estrutura, valores estocásticos, e análise no domínio das frequências.

Como medidas de texturas podemos citar:

- As abordagens estatísticas fazem uso da função de probabilidade da densidade de nível de cinza (pdf), que é computada, normalmente, como a probabilidade condicional de articulação de pares de *pixel* (nível de cinza), num determinado local da imagem. Como exemplos de aproximações estatísticas para a análise da textura pode-se citar: a matriz de co-ocorrência do nível de cinza (GLCM), sendo Haralick (1979), um dos pioneiros na sua utilização;
- Abordagens estruturais explorando a noção de textura primitiva: um conjunto de *pixels* são caracterizados pelos seus atributos, onde o *pixel* com o seu atributo nível de cinza é a mais simples forma de textura primitiva. Nessa abordagem se enquadram os métodos conhecidos por Gaussian Markov Random Fields (GMRFs), (Chellappa & Chatterjee, 1985);
- Abordagens estocásticas: assume que medidas estatísticas locais podem ser relacionadas a texturas, como por exemplo, a media local, o desvio padrão ou os métodos de características geométricas, “Statistical Geometrical Features” (SGF), (Chen et al., 1995);
- As análises no domínio das frequências são baseadas na análise de Fourier, que é conduzida no domínio das frequências espectrais da imagem (PSM – power spectrum method). Existe uma variedade de métodos desenvolvidos para separar texturas diferentes baseados no PSM, a citar os filtros de Gabor, (Bovik et al., 1990), e as transformadas wavelets, (Unser, 1995).

Dentre os métodos descritos, na extração de características de textura para uso em classificação de imagens, oriundas de sensoriamento remoto, pode-se citar os trabalhos de Christodoulou et al. (2001), Huawu Deng & Clausi, D.A. (2003) utilizando análise no domínio das frequências; Bin Tian et al. (1999), Peddle, D.R. & Franklin, S. E. (1991) e Musci, M. (2006), usando abordagens estatísticas e matrizes de co-ocorrência.

Veska et al. citados por Haralick & Shapiro (1992), ao fazerem estudos comparativos para determinar qual abordagem fornecia a melhor separação de feições espaciais concluíram que as abordagens no domínio espacial - medidas estatísticas simples, e probabilidades de nível de cinza e matriz de co-ocorrência - fornecem uma classificação mais precisa que a análise no domínio das frequências.

Também o trabalho de Walker (1997), enfatiza o uso de métodos estocásticos e estatísticos na classificação de imagens.

Os métodos utilizados na quantização das características de texturas usados no presente trabalho são melhor descritos no Capítulo 3.

2.3.2.1 Informações de Cor em Textura

Segundo Mäenpää & Pietikäinen (2004), o uso de textura colorida tem se tornado uma técnica popular na análise de imagens. Um dos primeiros a utilizá-la foi Rosenfeld et al. (1982), que utilizava interações espaciais entre as bandas espectrais da imagem.

Em algumas aproximações, somente a interação espacial entre as bandas espectrais é considerada. Por exemplo, Caelli & Reye (1993) propuseram um método que extrai características de canais espectrais usando filtros isotrópicos multiescalas. Técnicas conjuntas para obtenção de características das texturas envolvidas na imagem como o *correlograma*⁶ de cores, foram usadas por Huang (1997). Uma aproximação interessante em combinar cores e textura com histogramas foi recentemente publicada por Paschos & Petrou (2003).

Outra forma de análise de cores em texturas é a divisão do sinal colorido da imagem em componentes de luminância e crominância e processá-los separadamente. Um certo número de aproximações usando esse princípio tem sido propostas, Tan & Kittler (1993) usaram características de textura baseada na transformada discreta do co-seno (*DCT - discrete cosine transform*)⁷ de uma imagem em nível de cinza, enquanto medidas derivadas de histograma de cores foram usadas para descrever as cores. Dubuisson-Jolly e Gupta (2000) utilizaram um método para classificar imagens de sensoriamento remoto, cujos cálculos foram feitos no espaço de cores e textura de forma independente. Sendo a avaliação final obtida pelos classificadores de cor e textura.

De acordo com os resultados de Mäenpää & Pietikäinen (2004), quando existem variações de iluminação em diferentes partes da imagem, a abordagem de cor e textura

⁶ Representação gráfica das autocorrelações de um conjunto de dados.

⁷ Transformação na imagem, mudando o domínio de representação da mesma.

separadas produziu melhores resultados. Embora a análise conjunta de cor e textura tenha sido melhor em testes com imagens que não apresentam variações de luminância.

Ainda de acordo com Mäenpää, a diferença encontrada entre os dois métodos foi mínima, então foi escolhida a abordagem de análise de cor e textura em paralelo, para o presente trabalho, por se tratar de uma abordagem que não envolve cálculos custosos a nível computacional e pode-se trabalhar com somente a informação de textura, independentemente da informação de cor da imagem, com a utilização dos mesmos métodos empregados. Isso se torna útil para imagens oriundas de radar, que são normalmente monocromáticas.

2.3.3 Medição das Características

Uma vez definido o método de descrição de textura a ser utilizado a saber, SGF ou matriz de co-ocorrência, deve-se amostrar as regiões da imagem sob análise, sendo sua obtenção realizada pelo deslocamento de uma janela (*sliding window*) sobre a imagem, conforme mostra a Figura 5.

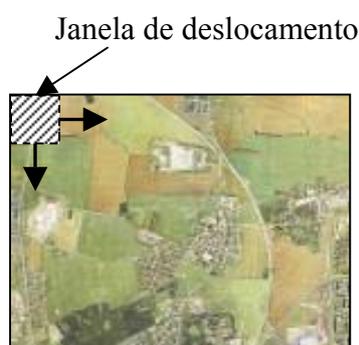


Figura 5 – Aplicação de uma janela de deslocamento na imagem em análise, para obtenção de medidas de características.

Ao deslocar-se essa janela, de dimensões $L \times C$ *pixels*, são coletadas medidas de características para análise posterior de sua classe temática (descrições das medidas serão contempladas no Capítulo 3). Os parâmetros de textura referentes à janela são então analisados, calculada a classe temática da janela e finalmente é marcado o *pixel* central dessa janela como pertencente a essa classe, a seguir, mover a janela *pixel* por *pixel*, repetindo esse mesmo procedimento (ver Figura 6).

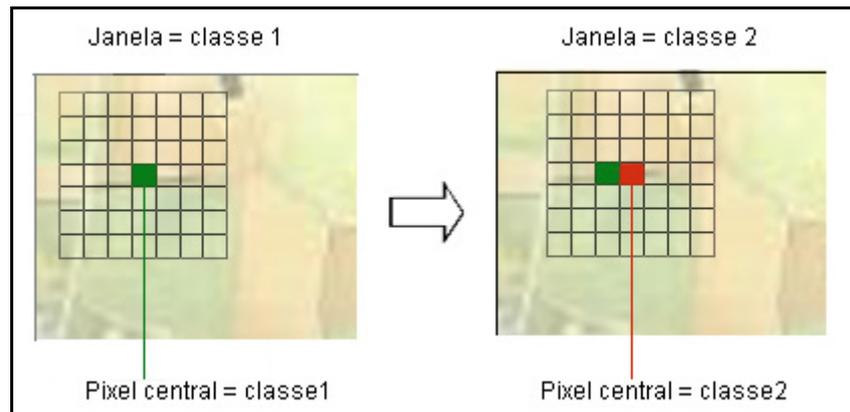


Figura 6 – Técnica de extração de características com janela deslizante – *slide*. A classe do *pixel* central é tomada a partir da classe da janela, a janela é então movida de um *pixel* e repetido o procedimento.

A coleta de informações de cores é efetuada tomando-se medidas de dispersão, como média e desvio padrão do conjunto de *pixels* da janela, semelhante trabalho foi efetuado por Briggs & Nellis (1991). Foram utilizadas as informações de matiz (H) e saturação (S) do sistema de cores HSV, sendo as informações de nível de cinza (V) utilizada para extrair-se características texturais da imagem.

A imagem original, colorida, é primeiramente decomposta em duas outras, uma contendo informações de cores, a saber H e S e outra contendo informações de níveis de cinza. O deslocamento da janela, como foi descrito acima, é composto na realidade pelo deslocamento de duas janelas simultâneas, ambas de igual tamanho, nas duas imagens criadas.

Se definirmos os operadores características de textura f_T e características de cor f_C , aplicados em uma janela \mathbf{J} , de tamanho $L \times C$ *pixels* - uma matriz $\mathbf{J}(L,C)$ - temos então para uma janela em particular:

- informação de níveis de cinza $\rightarrow f_T(\mathbf{J})$
- informação de matiz e saturação $\rightarrow f_C(\mathbf{J})$

Sendo f_T um operador de características de SGF ou co-ocorrência, dependendo do método empregado e f_C um operador que extrai média e desvio padrão.

O método empregado pode ser melhor visualizado na Figura 7, onde fica claro a utilização de duas imagens distintas, uma no espaço de cores, com informações de matiz e saturação e outra monocromática.

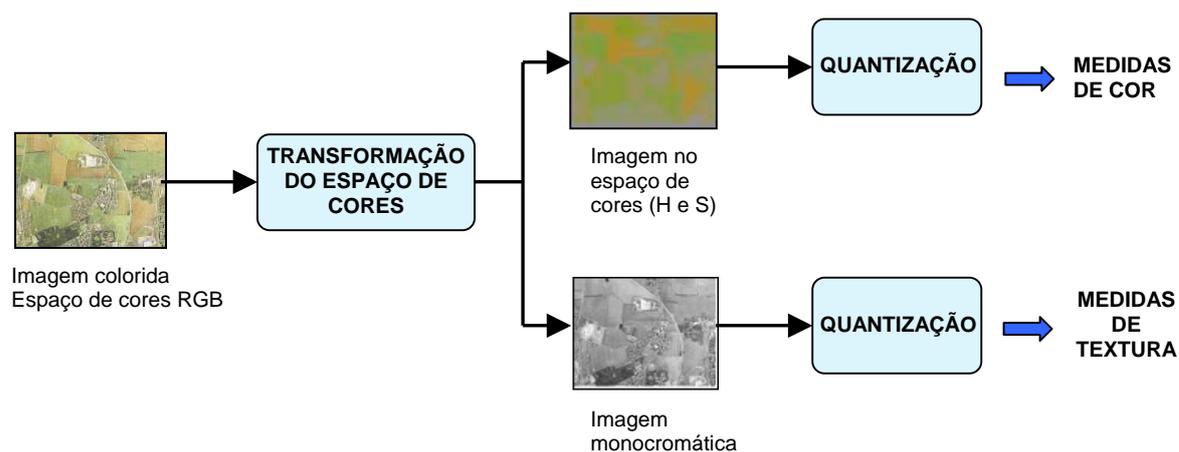


Figura 7 – Diagrama da transformação do espaço de cores RGB da imagem para a extração de características de cor e textura.

A natureza complexa de texturas em uma imagem tem sido muito investigada e vários esforços têm se concentrado em como capturar a variação local entre os *pixels* que a compõem. De acordo com Marceu (1990), mais de 90% de variação, na classificação de imagens usando texturas, pode ser encontrada apenas pela variação do tamanho da janela utilizada para extração de características. O tamanho da janela é de fundamental importância porque janelas de grandes dimensões são necessárias para capturar os vários tipos de textura e diminuir a variabilidade entre classes de textura diferentes. Entretanto, quanto maior a janela, maior será a sobreposição entre duas ou mais regiões, levando a erros de classificação entre essas regiões. O problema da determinação do tamanho da janela foi intensamente pesquisado por Glotfelty (1999) e um algoritmo foi sugerido para determinação automática da janela. Como a natureza do presente trabalho se concentra em analisar diferentes tipos de classificadores e métodos de extração de características de textura, o tamanho da janela foi determinado de forma empírica ao longo das experiências realizadas, ver Capítulo 5.

2.3.3.1 Aplicação de Heurística para Melhorar o Desempenho de Classificação

Freqüentemente erros de classificação devidos a diversos fatores, principalmente ao uso da janela de amostragem, ocorrem. Com isso um *pixel* pode ser incorretamente classificado e aparecer isolado dentro de uma região na qual não pertence. Pode-se conseguir uma melhora na classificação geral, pelo uso de uma heurística na imagem classificada resultante da seguinte forma: “Se na região de entorno de um *pixel* p , pertencente a uma classe c_1 , existirem seis ou mais *pixels* pertencentes a uma classe c_2 , diferente de c_1 , então fazer esse *pixel* p pertencer a essa classe c_2 ”.

2.4 Extração de Características

O problema em reconhecimento de padrões está na sua definição ou composição, já que definir um conjunto de características que o representa pode não ser uma tarefa trivial. A chave é escolher e extrair um conjunto finito de características que o represente totalmente e que possa ser passível de ser manuseado, (Bezdek & Pal, 1992). Entretanto, é comum que muitas dessas características possuam poucas ou nenhuma informação relevante para a discriminação das classes dos objetos envolvidos, ou seja, não possuem alta discriminação ou informação discriminatória (*discriminatory information*). Sendo assim essas informações são essenciais para o sucesso do processo de classificação. É de grande importância a correta seleção dessas características de forma a minimizar a sobreposição de áreas comuns entre as distribuições das características das classes envolvidas. Removendo características que não possuam informação discriminatória do conjunto de dados, podemos alcançar importantes vantagens, como: o subconjunto das características pode ser processado em menos tempo, simplesmente porque existem menos dados para o processamento e menores requisitos para armazenamento serão necessários; a redução do espaço de características pode aumentar a precisão das distribuições ditas classes-condicionadas.

Uma importante operação para a correta classificação dos dados obtidos pelas características obtidas é chamada de “maldição da dimensionalidade” (*curse of dimensionality*), esse termo foi criado por Bellman em 1961, e inclui um número de problemas que surgem com o aumento da dimensão dos dados. Este problema compromete, muitas vezes, a tarefa de reconhecimento de padrões. Em muitas aplicações, numerosas características são obtidas a fim de assegurar uma classificação acurada das classes desconhecidas. Algumas vezes, o número de características deve ser reduzido antes de usar um classificador, pois o método de classificação utilizado pode se tornar computacionalmente e/ou analiticamente intratável, (Poston & Marchette, 1998).

Para resolver problema de redução de dimensionalidade busca-se um espaço de dimensão menor que ainda represente a informação contida nestes dados. É razoável pensar assim, pois a quantidade de dados muitas vezes é confundida com a qualidade das relações existentes entre os mesmos. Na literatura existente, a redução de dimensionalidade está intimamente ligada à capacidade dos algoritmos em utilizar o espaço das características para promover uma boa classificação. Portanto, é útil ou mesmo necessário primeiro fazer um pré-processamento no sentido de reduzir a dimensão das características (dimensão dos dados) para um tamanho manuseável, ajustando tanto as informações originais quanto possíveis.

Pode-se pensar que um fenômeno representado por um espaço de padrões de alta dimensão, e assim complexo, pode realmente ser governado por umas poucas variáveis (espaço de características).

Os métodos de redução de dimensionalidade encontrados na literatura, (Webb, 2002 e Theodoridis & Koutroumbas, 2003), estudam diversas técnicas que permitem reduzir espaços de altas dimensões para espaços de dimensões menores.

Algumas das técnicas utilizadas para a redução de dimensionalidade podem ser descritas de duas formas diferentes. A primeira é identificar as variáveis que não contribuem para o processo de classificação, essas variáveis devem ser então desprezadas. O processo se caracteriza por desprezar um número d de características dentro do espaço de dimensão p (o número de características d precisa ser determinado). Esse processo é chamado de seleção de características. Em outras palavras, isso significa selecionar um subconjunto dentro de um grande conjunto de varias outras variáveis (ver Figura 8a).

A segunda aproximação consiste em encontrar uma transformação do espaço de dimensão p para um espaço de menor dimensão. Isso é chamado de extração de características (ver Figura 8b). Esta transformação pode ser uma combinação linear ou não-linear das variáveis originais e pode ser também supervisionada ou não-supervisionada.

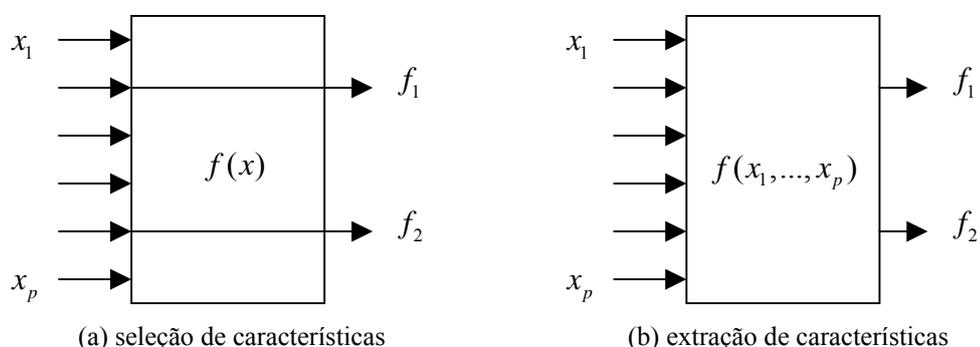


Figura 8 – Redução da dimensionalidade por (a) seleção de características (b) extração de características.

A aplicação de tais técnicas ocorre frequentemente em problemas de reconhecimento de imagem, voz, etc. Dentre as diversas técnicas, tem-se: análise de componentes principais, análise fatorial, análises discriminantes, curvas e superfícies principais e a *rede de Kohonen*⁸.

Algumas das técnicas utilizadas, nesta dissertação, para a redução de dimensionalidade são: análise de componentes principais e análises discriminantes.

⁸ rede de Kohonen – técnica principalmente usada para visualização dos dados de alta dimensão. Esta rede neural é também considerada uma forma de redução de dimensionalidade no sentido que ela aprende, de uma forma não supervisionada, um mapeamento entre um espaço l -dimensional e o espaço dos padrões.

2.4.1 Análise de Componentes Principais (PCA – *Principal Components Analysis*)

Também conhecida como uma variação da transformação de *Karhunen-Loeve*⁹, (Webb, 2002), é uma técnica de redução do número total de variáveis. Esta técnica é empregada quando existe redundância nos dados de uma amostra, onde redundância pode ser entendida como dados correlacionados, provavelmente em função de estarem medindo o mesmo evento. A existência de redundância é o que permite a redução no número de variáveis. Para verificar a existência ou não de redundâncias, a matriz de correlações entre as variáveis deve ser analisada. A técnica de PCA consiste em reescrever as coordenadas de um conjunto de dados em um outro sistema de eixos que seja mais conveniente para a análise desses dados. Estas novas coordenadas são o resultado da combinação linear das variáveis originais e são representadas sobre eixos ortogonais, sendo obtidas em ordem decrescente de variância. Portanto, a primeira componente principal detém mais informação sobre os dados do que a segunda componente principal que não detém informações contabilizadas anteriormente (na primeira componente principal) e assim sucessivamente. Em função da ortogonalidade dos eixos, as componentes principais não são correlacionadas. O número total de componentes principais é igual ao número total de variáveis originais e apresenta a mesma informação estatística que estas variáveis. Porém, este método permite a redução do número total de variáveis, pois frequentemente as primeiras componentes principais detêm mais de 90% da informação estatística dos dados originais, maiores informações podem ser vistas no Apêndice A.

Na Figura 9 pode-se visualizar o efeito da aplicação de PCA para a construção dos eixos ortogonais **1** e **2**, facilitando a separação dos conjuntos de variáveis **A** e **B**.

⁹ transformação de Karhunen-Loeve – processo de análise de componentes principais, na qual são levadas em consideração informações a respeito das classes envolvidas.

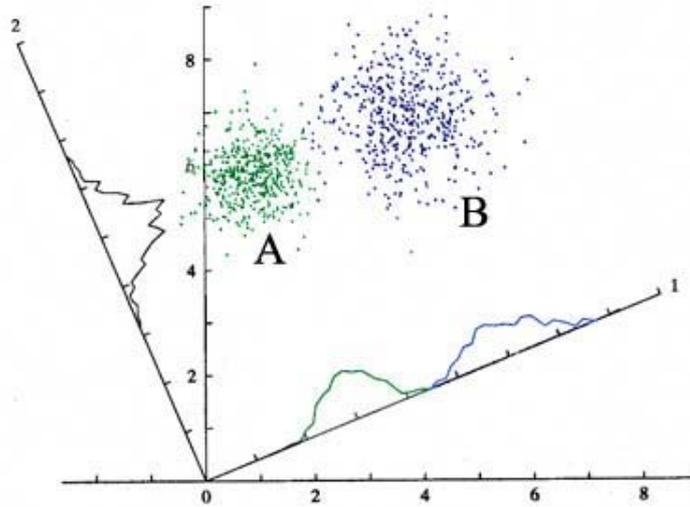


Figura 9 – Aplicação de PCA - Os dados são melhor visualizados na direção da dimensão dos eixos ortogonais **1** e **2** , mostrando claramente dois conjuntos diferentes **A** e **B**.

2.5 Pré-processamento de características

Técnica de transformar os dados de forma a serem melhor utilizados pelos processos subseqüentes. Classificadores baseados em redes neurais geralmente requerem que os dados de entrada $x \in \mathbb{R}$ estejam dentro dos limites $x \in \mathbb{R}' \subset \mathbb{R}$, isto é, entre -1 e 1, ou sejam valores inteiros, $x \in \mathbb{Z}$. Outros métodos incluem dados binários, $x \in \{0,1\}$ e modificações na distribuição estatística do conjunto de dados, como a normalização da média e variância para todos o elementos de dados. Qualquer que seja o método utilizado para o pré-processamento, é somente importante que seja mantida a ordem de relacionamento entre os dados, ou seja,

$$x_1 < x_2 \Rightarrow x'_1 < x'_2, \quad (3)$$

onde x' é o valor de x após o pré-processamento.

Como será visto no Capítulo 4, que trata de redes neurais, os classificadores propostos, baseados em redes neurais, tiveram os dados normalizados (média zero e desvio padrão um).

2.6 Seleção de Características

Processo de escolha, com a utilização de uma função de discriminação, J , de um subconjunto de tamanho d , X_d , a partir de um espaço de variáveis de dimensão p , denotadas por x_1, \dots, x_p . Desse modo procuramos o subconjunto \tilde{X}_d para o qual

$$J(\tilde{X}_d) = \max_{X \in X_d} J(X) \quad (4)$$

A descrição acima é somente uma simplificação da técnica de redução de dimensionalidade. A função de discriminação J é usualmente baseada em medidas de distancia ou desigualdade entre distribuições estatísticas, das variáveis do conjunto de características. Essas medidas de distancia requerem a definição de distancias entre os objetos envolvidos. Medidas estatísticas são apresentadas no Apêndice A.

2.6.1 Critérios para Seleção de Características

Na escolha de um conjunto de características ótimo, torna-se necessário um modo de mensurar a capacidade desse conjunto em discriminar com bastante eficiência as classes de objetos envolvidas. Isso pode ser conseguido definindo-se uma medida de separação de classes otimizada em relação a todas as classes envolvidas. Os conjuntos de características podem ser escolhidos essencialmente de duas formas possíveis:

1. Podemos construir um classificador para um conjunto reduzido de características e escolher aquelas que produzem um melhor resultado em um conjunto separado de teste/validação. Nesta aproximação, o conjunto de características é escolhido de forma a adaptar-se ao classificador. Um conjunto diferente de características pode resultar dependendo apenas do tipo de classificador envolvido.
2. A segunda aproximação consiste, em estimar a sobreposição entre as distribuições das classes de dados e escolher o conjunto de características para o qual essa sobreposição é mínima, ou seja, maximizar a separação. Isso é independente do tipo de classificador envolvido e possui a vantagem de que geralmente é de implementação simples, e a desvantagem de que as suposições feitas, em determinar as sobreposições das classes são muitas vezes rudimentares e podem resultar em discriminações não muito exatas.

Os métodos descritos diferem quanto à forma exata de escolha do conjunto de dados, mas ambos possuem a característica de designar uma figura de mérito para a escolha das melhores características a serem selecionadas.

Para o segundo método pode-se utilizar uma medida de seleção chamada de *distância de probabilidade*, que é a quantidade de sobreposição entre as distribuições de classes. Outras medidas quantificam a separação entre as distribuições, normalizadas pela dispersão da variância de cada distribuição, ver Apêndice A.

Medidas de distância de probabilidade são alguns dos métodos comumente aplicados em medidas de separação e possibilitam quantificar as distancias entre duas funções de densidade de probabilidade (pdf). Quanto maior essas distâncias, menor será a sobreposição entre as densidades e menor o erro na classificação. Algumas medidas comumente utilizadas são *Chernoff*, *Divergence*, *Mahalanobis* e *Bhattacharyya*, sendo esta ultima a mais utilizada, uma descrição detalhada dessas medidas pode ser vista em Webb (2002).

Conforme mostrado no Apêndice A, para se fazer uso dessa medida deve-se supor uma distribuição gaussiana¹⁰ dos dados. Segundo Gonzáles & Woods (2000), as propriedades estatísticas das classes de padrões em um problema são freqüentemente desconhecidas ou não podem ser estimadas. Na pratica, tais problemas de decisão teórica são melhor tratados por métodos que levem diretamente as funções de decisão através do treinamento.

No presente trabalho, nada pode ser afirmado a respeito dos dados provenientes da extração de características das imagens utilizadas, ou seja, não se pode dizer que os dados sejam do tipo gaussiano, o que dificulta a sua utilização com os métodos supracitados, a menos que se utilize uma transformação dos dados para o tipo gaussiano, por intermédio de técnicas como *Ladder of Powers*¹¹ ou *Box-cox*¹², como no trabalho de Walker (1997). A utilização dessas técnicas requer uma análise estatística multivariada de todos os dados extraídos, existindo portanto a possibilidade de auferir erros posteriores na classificação. Além disso, de acordo com a literatura, Webb (2002), Theodoridis & Koutroumbas (2003) e Sing-Tzr-Bow (2002), somente funções de discriminação lineares ou quadráticas são mais

¹⁰ Distribuição de probabilidade que se caracteriza por uma curva em forma de sino em torno do valor médio.

¹¹ Método utilizado para normalizar uma seqüência de dados estatísticos com somente uma variável; Velleman, P. F. & Hoaglin, D. C. (1981) "Applications, Basics, and Computing of Exploratory Data Analysis, Duxbury Press, Boston, Mass.

¹² Estima uma aproximação normal para dados estatísticos multivariados, utilizando o método de probabilidade máxima; Box, G. E. P. & Cox, D. R. (1964), "An analysis of transformations", Journal of the Royal Statistical Society, 26, 211-252.

sensíveis ao tipo de distribuição gaussiana, no caso dessa dissertação, serão utilizadas redes neurais para classificação dos dados, tendo então, pouca importância o tipo de distribuição envolvida. Sendo então utilizada a primeira aproximação citada a saber: construção de um classificador para um conjunto reduzido de amostras.

2.6.2 Algoritmos de Busca para Seleção de Características

Uma vez definido o tipo de função de discriminação, a ser utilizada para a seleção das melhores características, dentre todas envolvidas, resta escolher o tipo de algoritmo para essa escolha.

O problema de seleção de características está em escolher o melhor subconjunto de tamanho d a partir de um espaço de variáveis de dimensão p .

Essa escolha é feita construindo-se um conjunto de d características de forma incremental, ou seja, começando com um conjunto vazio e incluindo elementos (método *bottom-up*), ou a partir de um conjunto contendo todas as características disponíveis e removendo aquelas redundantes (método *top-down*).

Se X_d representa um conjunto de características ou variáveis, então pelo método *bottom-up*, o melhor conjunto, \tilde{X}_d , em cada passo do algoritmo é aquele para o qual o critério de seleção de características possui o valor Máximo, segundo a equação (4).

O conjunto X_d de todas as características em um determinado passo do algoritmo, é determinado pelo conjunto do passo anterior. Isso significa que a dimensão do conjunto utilizada em determinada etapa do algoritmo é utilizada como ponto de partida para determinar o próximo conjunto de valores no passo seguinte. Isso não significa que os conjuntos sejam necessariamente aninhados ($\tilde{X}_d \subset \tilde{X}_{d+1}$), embora eles possam ser.

Um algoritmo de busca ótima que não envolve uma busca exaustiva pelo melhor conjunto de características é o método “*branch and bound*”¹³. Esse algoritmo é um procedimento *top-down*, o qual começa com um conjunto de p variáveis e constrói uma árvore de busca excluindo sucessivamente as variáveis envolvidas. Para o correto funcionamento desse algoritmo, deve ser observado uma importante propriedade de seleção de características, propriedade esta definida da seguinte forma: dados dois conjuntos de variáveis X e Y ,

$$X \subset Y \Rightarrow J(X) < J(Y) \quad (5)$$

¹³ Algoritmo proposto por A. H. Land e A. G. Doig em 1960.

Isto é, se X é um subconjunto de Y , então a avaliação do critério de seleção de características, seguindo uma determinada função, desse subconjunto deve ser menor do que aquela na qual esse conjunto está inserido. Isso é chamado de propriedade monotônica, a qual deve ser assegurada para a utilização do algoritmo mencionado.

Entretanto mesmo quando essa propriedade for assegurada, esse método pode ser proibitivo em termos computacionais para um grande conjunto de características a serem selecionadas.

Outros algoritmos de busca produzem um resultado aceitável, inclusive sem a necessidade de satisfazer a propriedade monotônica descrita. Esses algoritmos produzem um resultado sub-ótimo, efetuando uma busca em apenas um subconjunto dentre todos os possíveis. Exemplos desses algoritmos, vistos em detalhe em Webb (2002), são:

- *sequential forward selection* (SFS), é uma busca do tipo *bottom-up*, a qual adiciona novas características a cada passo, até que o conjunto final de características seja alcançado. A maior desvantagem desse método, é que uma vez selecionadas, as características não podem ser retiradas do conjunto final.
- *sequential backward selection* (SBF), ao contrario do SFS, é uma busca do tipo *top-down*. As variáveis são excluídas uma após a outra, até que somente o conjunto de características desejadas seja alcançado. O ponto fraco desse algoritmo é que uma vez retiradas, as características não podem ser adicionadas novamente.
- *plus l – take away r selection*, este algoritmo possibilita voltar atrás no processo de seleção de características. Cada passo do algoritmo adiciona l características no subconjunto, aplicando o método SFS l vezes, simultaneamente remove r características aplicando o método SBS r vezes.

Segundo Kittler (1978), o método *plus l – take away r* demonstra ótimos resultados e é computacionalmente satisfatório para a maioria das aplicações. Por essa razão esse algoritmo foi escolhido para a seleção de características. Sendo a função de discriminação construída a partir de um classificador que utiliza redes neurais, ver capítulo 4. Na Figura 10 pode ser visto detalhes do módulo de seleção de características, que utiliza o algoritmo *plus l – take away r* descrito acima, cuja função de discriminação é baseada em um classificador que utiliza uma rede neural, do mesmo tipo utilizada no módulo classificador principal, usado para classificar a imagem em análise.

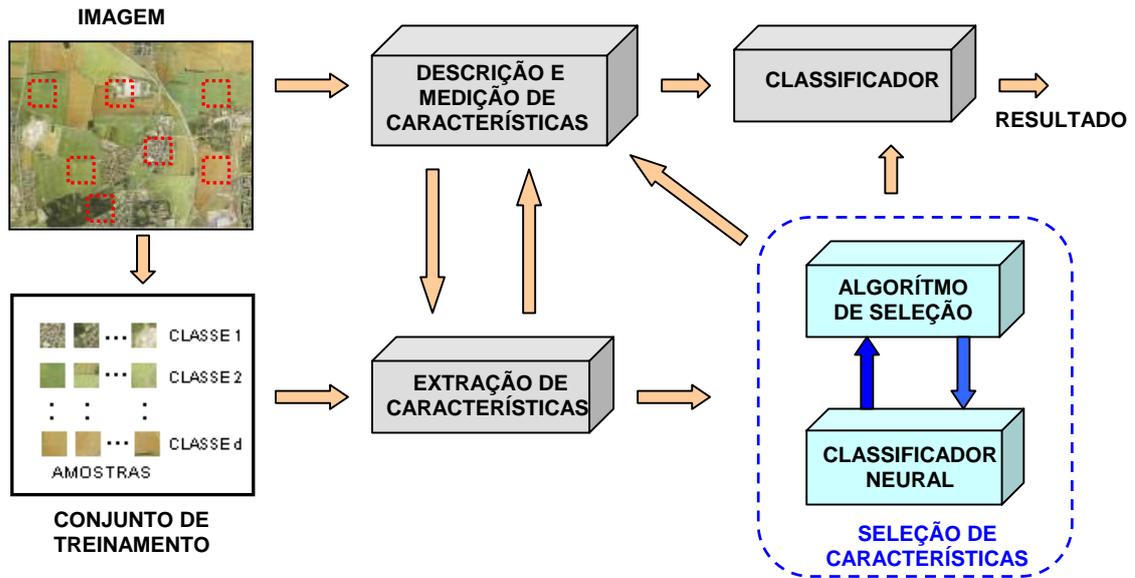


Figura 10 – Módulo de seleção de características, que utiliza uma função de discriminação baseada em um classificador neural e um algoritmo de seleção sub-ótimo (plus 1 take away r).

2.7 Classificação

O processo de classificação usa a informação representada pelos números digitais das características selecionadas da imagem e tenta classificar cada *pixel* individual baseado em informações dessas características. O objetivo é designar todos *pixels* na imagem em classes particulares ou temas (e.g. água, floresta de coníferas, deserto, milho, trigo, etc.). A imagem classificada resultante é constituída de um mosaico de *pixels*, cada um pertencente a um tema particular, isso é essencialmente um mapa temático da imagem original.

Quando se fala sobre classes, é necessário distinguir entre classes de informação e classes espectrais. Classes de Informação são aquelas categorias de interesse que o analista está realmente tentando identificar na imagem, tal como espécies diferentes de colheitas, tipos de floresta diferentes ou espécies de árvores, unidades geológicas diferentes ou tipos de rocha, etc. Classes espectrais são agrupamentos de *pixels* uniformes (ou bem parecidos) em relação a seus valores de brilho nos diferentes canais do espectro de dados. O objetivo está em casar as classes espectrais dos dados nas classes de informação de interesse.

Procedimentos comuns de classificação podem ser divididos dentro de duas subdivisões baseadas no método utilizado: *Classificação supervisionada* e *classificação não-supervisionada*.

2.7.1 Classificação não-supervisionada

As classes espectrais ou conjuntos de *pixels* com características semelhantes são agrupadas, baseado somente na informação numérica dos dados, e então são casadas por um analista para classes de informação (se possíveis). Programas, chamados *clustering algorithms*, são utilizados para determinar grupos naturais (estatísticos) ou estruturas nos dados. Usualmente, o analista especifica quantos grupos ou *clusters* deverão ser esperados nos dados. A classificação não-supervisionada não é realizada completamente sem a intervenção humana. Entretanto, esse tipo de classificação não começa com um predeterminado conjunto de classes como em uma classificação supervisionada.

Algoritmos comuns que se utilizam dessa técnica são o K-means, ISODATA e o C-means, sendo o último mais conhecido por uma variação chamada Fuzzy C-means ou FCN.

O K-means e o ISODATA funcionam de forma semelhante, sendo que para o funcionamento do K-means deve-se conhecer a princípio o número total de classes envolvidas, enquanto que no ISODATA isso não é necessário.

Ao contrario da abordagem clássica na qual cada elemento pertence totalmente a uma única classe, a abordagem nebulosa¹⁴ (*fuzzy*) feita pelo algoritmo FCM permite que um elemento possa pertencer a várias classes, com diferentes valores de pertinência, permitindo então uma melhor classificação em elementos que possam pertencer a diferentes classes. Na sua utilização deve-se conhecer o número de classes envolvidas na classificação.

Os trabalhos de Kim et al. (1997), Martin (2003), Andrey & Tarroux (1998), Deng & Manjunath (2001) e O'Callaghan & Bull (2005), fazem uso de texturas e classificação não supervisionada em imagens.

2.7.2 Classificação supervisionada

Em uma classificação supervisionada, o analista identifica na imagem amostras homogêneas representativas de diferentes tipos de áreas (informação de classes) de interesse. Essas amostras são referidas como áreas de treinamento, ou conjunto de treino. A seleção de áreas de treinamento apropriadas é baseada no conhecimento do especialista da geografia representada na imagem. Assim, o analista está "supervisionando" a categoria de um conjunto

¹⁴ Lógica que permite representar valores de pertinência (grau de verdade) intermediários entre os valores de verdadeiro e falso da lógica clássica (bivalente).

de classes específicas. A informação numérica contida em todos os *pixels*, compreendidos nessas áreas, é utilizada para treinar o programa em reconhecer áreas parecidas para cada classe. O computador usa um programa especial ou algoritmo (de que há várias variações), para determinar as "assinaturas" numéricas para cada classe de treinamento. Uma vez que o computador tenha determinado as assinaturas para cada classe, cada *pixel* na imagem é comparado com essas assinaturas e rotuladas como a classe que mais se aproxima digitalmente, ver exemplo dado pela Figura 11. Assim, em uma classificação supervisionada, primeiro são identificadas as classes de informação, que então são utilizadas para determinar as classes espectrais que as representam.

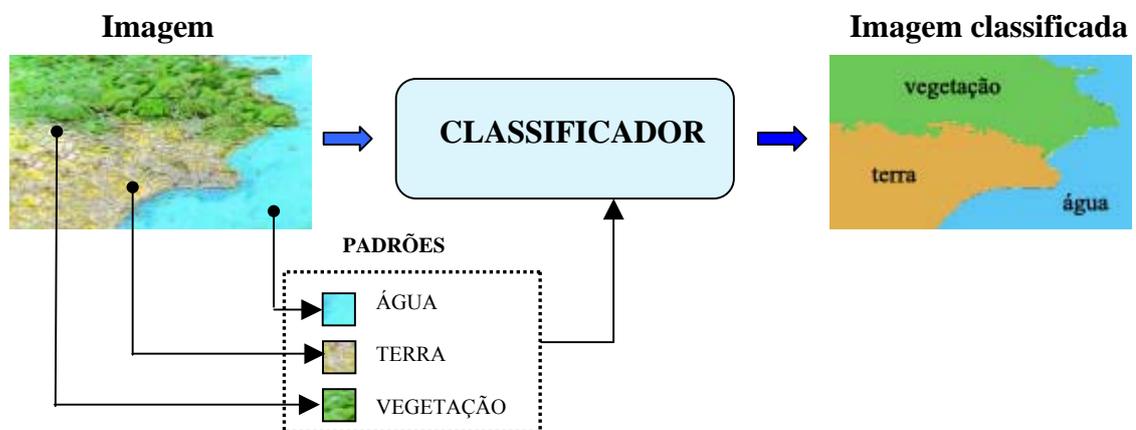


Figura 11 – Classificação de uma imagem utilizando padrões de textura e método supervisionado. Os padrões são selecionados, pelo analista, da imagem original e utilizados para treinar um classificador, resultando em uma imagem classificada com as classes previamente definidas.

A classificação supervisionada pode ser realizada com a utilização de algoritmos estatísticos ou com o uso de Redes Neurais Artificiais – RNAs. Algoritmos estatísticos usam parâmetros derivados das amostras de dados, de modo a estabelecer e identificar as classes no processo de classificação. Parâmetros, tais como valores mínimos e máximos, ou valores médios, ou desvios padrões, são utilizados como elementos para definir as classes existentes em uma imagem. Os algoritmos neurais utilizam diretamente as amostras de dados como conjunto de dados de treinamento de sua rede neural, possibilitando a identificação das classes por padrões de reconhecimento. Assim, os métodos estatísticos são chamados paramétricos, pois não utilizam diretamente o conjunto de dados de entrada, mas parâmetros

estatísticos derivados desse conjunto de informações. Em contraposição, os métodos neurais são chamados não paramétricos.

Os classificadores estatísticos não possuem um grande poder de generalização, ocasionando com isso uma grande quantidade de *pixels* não classificados ou classificados incorretamente. Os tipos mais comuns de classificadores estatísticos supervisionados utilizam os métodos do paralelepípedo, da distância mínima e da máxima verossimilhança.

Uma abordagem promissora e amplamente utilizada na classificação supervisionada é a que usa redes neurais, ver Capítulo 4.

Com a capacidade intrínseca de aprender, as redes neurais surgiram como uma ferramenta poderosa na área de reconhecimento de padrões. Sua capacidade de fazer suposições mais delicadas, a respeito da distribuição dos dados de entrada, do que métodos estatísticos tradicionais e a capacidade de formar fronteiras de decisão, altamente não-lineares, no espaço de características, levaram ao seu crescente uso, como bem afirmou Souza (1999). Uma rede neural, também conhecida como rede conexionista, é definida em Cichoki e Unbehauen (1993), como um sistema de processamento de sinal ou informação, composto por um grande número de elementos simples de processamento, chamados neurônios artificiais, ou simplesmente nós, que são interconectados por elos diretos chamados de conexões, que colaboram para realizar um processamento paralelo distribuído para realizar uma tarefa computacional desejada.

Na maioria das aplicações de RP, usando-se redes neurais, são estabelecidas conexões entre os valores de todas as características que definem os padrões, com camadas intermediárias de neurônios e a todas as classes objetivos, que são consideradas como a camada de saída. O treinamento da rede é realizado corrigindo os pesos nas conexões, para estabelecerem-se as relações entre as características e classes que promovam a melhor discriminação possível, entre os padrões de classes diferentes. Assim, ao ser apresentado à rede um novo padrão, esta indicará a classe que o representa melhor na camada de saída.

Na Figura 12 pode ser visualizado um exemplo de rede neural artificial, utilizada para classificação de texturas em uma imagem.

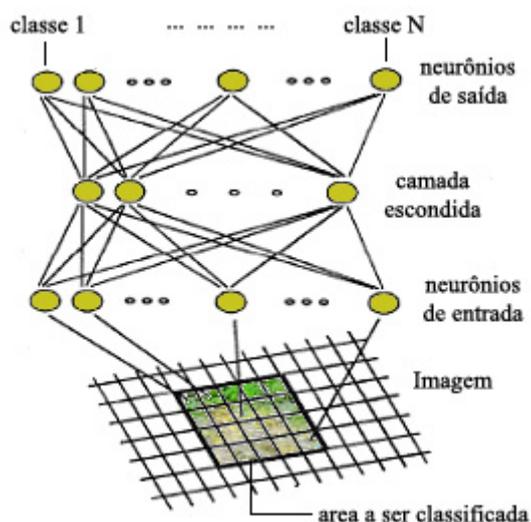


Figura 12 – Rede neural para classificação de texturas.

A modelagem de uma RNA pode ser dividida em 3 níveis: o neurônio artificial, a topologia ou arquitetura da rede e o processo ou algoritmo de aprendizado. Maiores detalhes são expostos no Capítulo 4, que define a utilização e construção das RNAs usadas nessa dissertação.

A utilização de Redes Neurais Artificiais com o propósito de classificação supervisionada pode, então, se constituir em uma alternativa vantajosa aos classificadores baseados em conceitos estatísticos, uma vez que nenhuma hipótese prévia sobre a distribuição dos dados a serem classificados é exigida, não restringindo a possibilidade de incorporação de dados não espectrais na classificação. Possuem ainda a capacidade de separar dados provenientes de superfícies de decisão complexas, ver Figura 13.

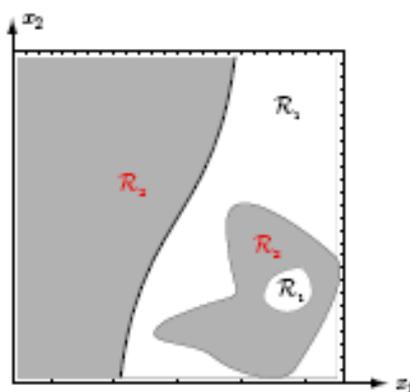


Figura 13 – Regiões de decisão complexas (com duas classes \mathcal{R}_1 e \mathcal{R}_2 não conectadas), que podem ser solucionadas por uma rede neural.

Existem vários modelos de redes neurais artificiais desenvolvidos para as mais diversas aplicações. Sendo atualmente as redes mais usadas em RP são: *perceptron* multicamadas (MLP – *Multi-Layer Perceptron*) com algoritmo de aprendizagem *backpropagation*, a rede de função de base radial (RBF – *Radial Basis Function*), a rede LVQ (Learning Vector Quantization), e os mapas topológicos de Kohonen (SOM - *Self-Organizing Map*), sendo que a última é utilizada para classificação do tipo não supervisionada, em alguns trabalhos é utilizada também a rede probabilística (PNN – *Probabilistic Neural Network*), sendo esta uma variação da RBF. Variações e combinações dessas redes são utilizadas em alguns trabalhos, a citar os de Arrowsmith et al. (1999) e Pitiot et al. (2002).

Segundo Gonzáles & Woods (2000), muitas vezes, as propriedades estatísticas de uma classe não podem ser estimadas, de modo que os problemas de teoria de decisão são mais bem trabalhados pelos métodos que produzem diretamente as funções de decisão através dos dados de treinamento, sem levar em conta qualquer informação probabilística das classes de interesse. Sendo assim, a escolha por um classificador baseado em RNAs torna-se uma escolha natural para utilização em imagens para fins de geoprocessamento, trabalhos semelhantes utilizando diversos tipos de RNAs podem ser vistos em Rosenfeld et al. (1982), Andrey & Tarroux (1998), Connors & Harlow (1978), Dubuisson & Gupta (2000), Galloway (1995), Glotfelty (1999) e Mäenpää & Pietikäinen (2004).

No presente trabalho será feita uma avaliação de desempenho dos modelos de RNAs MLP e PNN na classificação de imagens com texturas. A escolha por esses modelos leva em consideração os baixos desempenhos das redes SOM e LVQ em comparação com as redes MLP e RBF conforme estudos comparativos de Visa (1990), Christodoulou (2001), Yiu et al (1999) e Arrowsmith et al. (1992)

2.7.3 Estimativa de Erro

Avaliação de desempenho de um classificador é uma tarefa que quantifica como o sistema classifica bem os objetos, e é geralmente expressa em termos de erro de classificação, isto é, a taxa em que um objeto é alocado para uma classe diferente da que pertence.

Um dos parâmetros mais utilizados e eficientes na determinação da exatidão é o *coeficiente Kappa* (Apêndice A), o qual é obtido mediante a adoção de uma referência para a comparação dos mapeamentos produzidos. Neste caso, a análise da exatidão dos mapeamentos é obtida por meio da *matriz de confusão* ou *matriz de erro* (Apêndice A), sendo posteriormente calculado o coeficiente *Kappa* de concordância, (Congalton et al. 1993).

Capítulo 3

QUANTIZAÇÃO DE TEXTURA

3.1 Introdução

Como visto no Capítulo 2 que trata, entre outras, das diferentes formas de caracterização de textura, como abordagens estatísticas, estruturais, estocásticas e no domínio das frequências. Não há uma definição geral para textura, tal dificuldade é refletida pelo grande número de métodos de análise de texturas encontrados na literatura. Haralick et al.(1973), definiram textura a partir de medidas de uniformidade, densidade, aspereza, regularidade e intensidade, dentre outras características de imagem.

A extração de características de texturas a partir de matriz de co-ocorrência está entre os métodos mais utilizados da abordagem estatística para análise de textura. Métodos mais recentes foram sugeridos por Chen et al. (1995), que utilizam uma abordagem orientada a distribuição estatística de formas geométricas.

Este capítulo apresenta uma descrição de dois métodos de medidas para segmentação de imagens, a saber: matriz de co-ocorrência e características geométricas.

3.2 Matriz de Co-ocorrência

Co-ocorrência é uma medida de frequência relativa, ou probabilidade condicional, da ocorrência de propriedades de duas imagens, dentro de algumas limitações, a partir do domínio da imagem. Propriedades da imagem são: intensidade de *pixels*, variação da intensidade, medidas de gradiente, etc. Essas propriedades podem ser mensuradas em um mesmo espaço de características (magnitude e orientação) e outras definições de vizinhança. Por exemplo, matriz de co-ocorrência de níveis de cinza (GLCM – *Gray Level Co-occurrence Matrix*) é uma tabulação de quantas combinações diferentes de valores de intensidade dos *pixels* (níveis de cinza) ocorrem em uma imagem. A idéia principal é descrever textura através

de um conjunto de características para as ocorrências de cada nível de cinza nos *pixels* da imagem considerando múltiplas direções. Antes da aplicação da matriz de co-ocorrência, as imagens precisam passar por um processo de pré-processamento usando uma re-quantização ou uma equalização do histograma, ver Capítulo 5. Esse processo assegura domínios consistentes de co-ocorrência independente de propriedades estatísticas de primeira ordem de uma imagem como, média, máxima ou mínima intensidade de *pixels*.

Outros métodos comumente utilizados para caracterização de texturas, baseados em matriz de co-ocorrência, são: GLRLM (Gray Level Run Length Matrix) definida por Galloway (1995), GLVM (Gray Level Variance Matrix) definida por Yogesan (1994), SFM (Statistical Feature Matrix) por Wu & Chen (1992), NGLDM (Neighbouring Gray Level Dependence Matrix) por Sun & Wee (1983) e finalmente GCM (Generalised Co-occurrence Matrices) por Davis et al. (1979).

3.2.1 Matriz de Co-ocorrência de Níveis de Cinza - GLCM

Essa técnica foi inicialmente introduzida por Haralick et al. (1973). Primeiramente eles caracterizaram textura como uma média espacial de uma relação entre os tons de cinza de uma imagem. O propósito foi demonstrar que medidas estatísticas de segunda ordem são suficientes para discriminação de texturas.

GLCM estão entre as principais técnicas para descrever texturas utilizando estatísticas de segunda ordem, como a probabilidade de dois *pixels* terem um determinado nível de cinza ocorrendo com um relacionamento espacial particular.

Segundo Schwartz (2003), a matriz de co-ocorrência é uma tabulação de quantas combinações diferentes de valores de intensidade dos *pixels* (níveis de cinza) ocorrem em uma imagem. A idéia principal da matriz de co-ocorrência é descrever textura através de um conjunto de características para as ocorrências de cada nível de cinza nos *pixels* da imagem considerando múltiplas direções.

Na determinação da GLCM, podemos modelar uma imagem em escala de cinza em um domínio $D \subset \mathbb{Z}^2$ de N_g níveis de cinza como uma função em duas dimensões (2D) $I : D \rightarrow G$, onde $G = \{1, \dots, N_g\}$. A GLCM $P(i, j | d, \theta)$ é uma estimativa da função densidade de probabilidade condicional de segunda ordem, na qual dois *pixels*, separados pelo deslocamento d, θ , tendo intensidades i e j , onde $i, j \in G$:

$$P(i, j | d, \theta) = \frac{\#\{k, l \in D | I(k) = i, I(l) = j, \|k - l\| = d, \angle(k - l) = \theta\}}{\#\{m, n \in D | \|m - n\| = d, \angle(m - n) = \theta\}}, \quad (6)$$

onde k, l, m, n são localizações válidas de *pixels* na imagem. Devido a natureza discreta das intensidades da imagem, P é de fato, uma densidade discreta ao invés de contínua. Sendo uma densidade de probabilidade, para qualquer d, θ dado, temos:

$$\sum_{i,j} P(i, j | d, \theta) = 1, \quad (7)$$

e

$$\forall i, j \in G, \quad 0 \leq P(i, j | d, \theta) \leq 1. \quad (8)$$

Co-ocorrência na sua forma geral, pode ser especificada por uma matriz de freqüências relativas $P(i, j | d, \theta)$, na qual dois elementos de textura vizinhos, separados por uma distancia d em uma orientação θ ocorrem na imagem, um com propriedade i e outro com propriedade j , (Schwartz, 2003). A Figura 14 mostra a disposição espacial dos *pixels* para o cálculo da matriz de co-ocorrência.

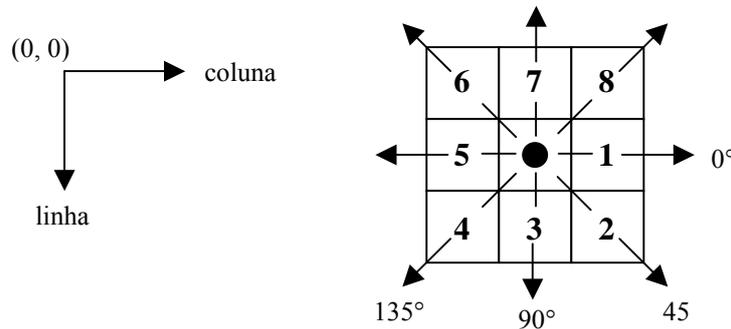


Figura 14 – Disposição espacial dos *pixels*.

Um exemplo, citado por Walker (1997), a Figura 15a mostra uma imagem em escala de cinza contendo variações de intensidade de 1 a 4. A GLCM correspondente, calculada para um par de *pixels* na horizontal deslocados de 1 *pixel* ($d = 1, \theta = 0^\circ$), é mostrado na Figura 15b. Por simplicidade é mostrado apenas a freqüência de ocorrência dos pares de *pixels* (numerador da equação (6)), ao invés da probabilidade normalizada.

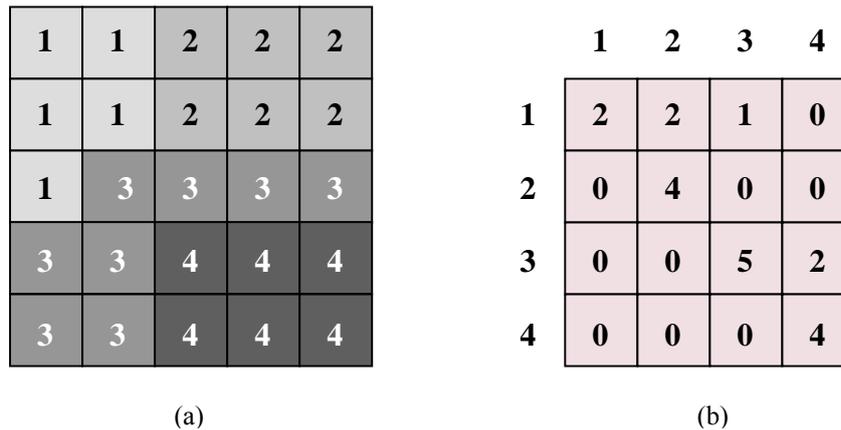


Figura 15 – Um exemplo do cálculo da GLCM em uma imagem com escala de cinza. (a) Imagem contendo quatro níveis de cinza; (b) a GLCM de (a) com $d = 1, \theta = 0^\circ$. O elemento $P(1,2)$ da matriz de co-ocorrência é determinado primeiramente procurando na imagem por todos os *pixels* que possuem intensidades iguais a 1. Uma vez encontrados esses elementos, conta-se todos os que possuem vizinhos com intensidade igual a 2, dispostos a uma distância de 1 *pixel* para direita ($d = 1, \theta = 0^\circ$). Existem dois *pixels* que satisfazem essa condição na imagem, então $P(1,2) = 2$.

Algumas considerações são necessárias, de forma a limitar o número de níveis de cinza N_g da imagem, a distancia d entre os *pixels* e a orientação θ . As imagens são geralmente re-quantizadas para 16 ou 32 níveis de cinza usando algumas técnicas de pré-processamento, ver seção 2. A orientação θ é geralmente limitada a quatro ângulos em intervalos de 45° , sem distinção entre os ângulos opostos, isto é, são utilizadas as orientação para $\theta = (0^\circ, 45^\circ, 90, 135^\circ)$. Algumas vezes são utilizadas uma média das matrizes de todos os valores de θ . A distancia d é escolhida de acordo com a rugosidade das imagens a serem analisadas, e geralmente varia de 1 até um valor máximo dependendo da textura da imagem, (Gool et al., 1990).

Na Figura 16 são mostrados 4 diferentes tipos de imagens de texturas. Todas foram quantizadas para 16 níveis de cinza e usadas para formar os histogramas de GLCM. Os parâmetros utilizados foram $d = 1$ e $\theta = 135^\circ$.

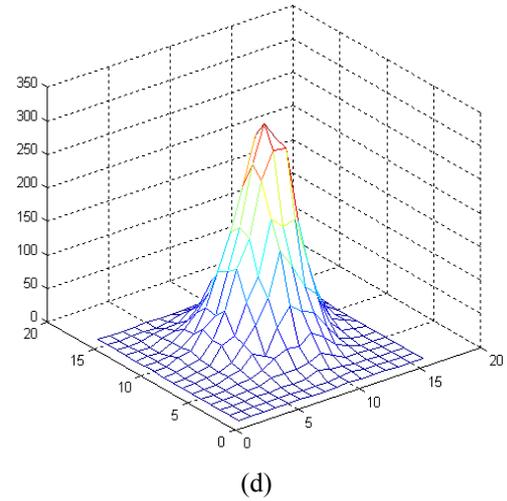
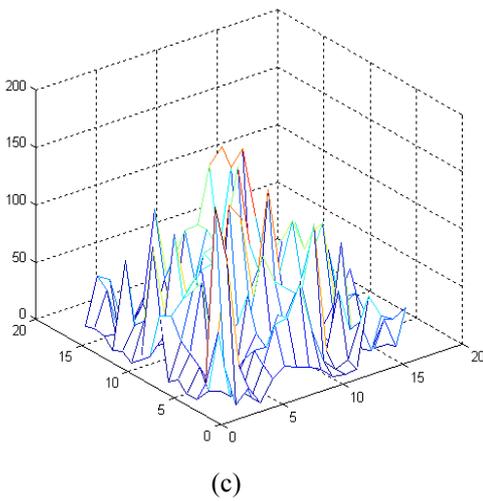
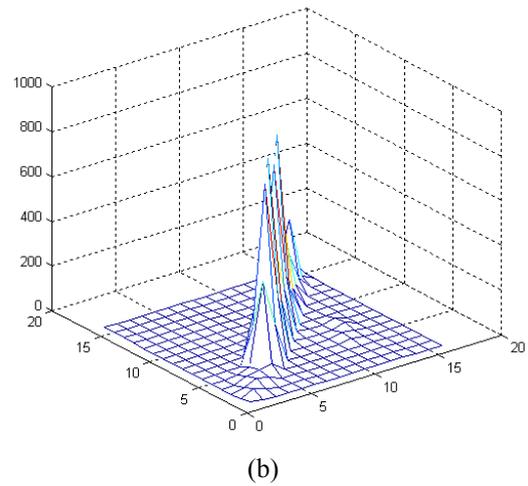
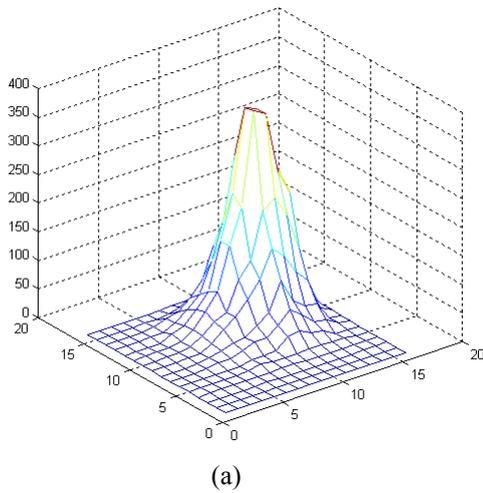
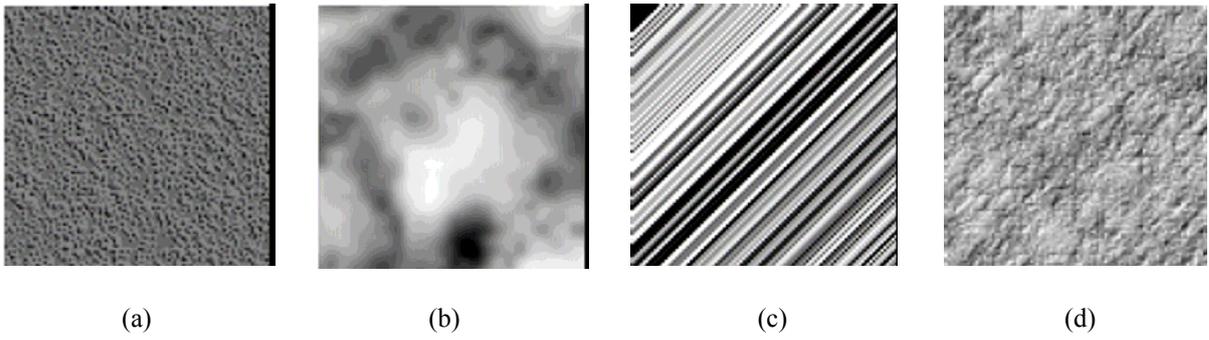


Figura 16 – Histogramas de GLCM para quatro tipos de imagens de textura. Todas as imagens possuem 96×96 pixels e foram quantizadas com 16 níveis de cinza antes da aplicação da matriz de co-ocorrência ($d = 1$ e $\theta = 135^\circ$). Fonte: Gonzáles e Woods (2002).

Weszka et al. (1976) verificou que se uma textura possui granularidade¹⁵ grossa e a distância d usada para o cálculo da GLCM for pequena, comparada ao tamanho dos elementos de textura, pares de *pixels* a uma distância d devem ter níveis de cinza similares. Isto significa que, altos valores na matriz $P(i, j | d, \theta)$ devem estar concentrados ou próximos da diagonal principal. De modo oposto, para uma textura de granularidade fina, se d é proporcional ao tamanho dos elementos de textura, então os níveis de cinza, de pontos separados por d , devem ser bem diferentes, com isso os valores em $P(i, j | d, \theta)$ devem estar uniformemente afastados. Similarmente se uma textura é direcionada, isto é, áspera em uma direção, o grau de dispersão dos valores da diagonal principal em $P(i, j | d, \theta)$, devem variar de acordo com a orientação θ .

Aplicações típicas de GLCM como definida por Haralick et al. (1973), envolve a extração de características escalares secundárias da matriz de co-ocorrência. Haralick et al. Definiu 14 funções características que foram estendidas por Connors, Trivedi & Harlow (1984) e outros. Na Tabela 1, é mostrada uma lista das funções de características de segunda ordem de GLCM mais utilizadas.

¹⁵ Nível de detalhe da textura.

Tabela 1 - Funções características de GLCM mais utilizadas

Funções	Equações
Energia	$F1 = \sum_{i,j} P(i, j)^2;$
Entropia	$F2 = -\sum_{i,j} P(i, j) \log P(i, j);$
Homogeneidade	$F3 = \sum_{i,j} \frac{1}{1 + (i - j)^2} P(i, j);$
Inércia	$F4 = \sum_{i,j} (i - j)^2 P(i, j);$
Correlação	$F5 = -\sum_{i,j} \frac{(i - \mu_x)(j - \mu_y)}{\sigma_x \sigma_y} P(i, j);$
Shade	$F6 = \sum_{i,j} (i + j - \mu_x - \mu_y)^3 P(i, j);$
Prominance	$F7 = \sum_{i,j} (i + j - \mu_x - \mu_y)^4 P(i, j);$
Variância	$F8 = \sum_{i,j} (i - \mu_x)^2 P(i, j);$
Soma da Média	$F9 = \sum_{i=2}^{2N_g} iP_{x+y}(i);$
Soma da Entropia	$F10 = -\sum_{i=2}^{2N_g} P_{x+y}(i) \log P_{x+y}(i);$
Soma da variância	$F11 = \sum_{i=2}^{2N_g} (i - F9)^2 P_{x+y}(i);$
Diferença da Média	$F12 = \sum_{i=0}^{N_g-1} iP_{x-y}(i);$
Diferença da Entropia	$F13 = \sum_{i=0}^{N_g-1} -P_{x-y}(i) \log P_{x-y}(i);$
Diferença da variância	$F14 = \sum_{i=0}^{N_g-1} (i - F12)^2 P_{x-y}(i);$
Medidas de Informação	$F15 = \frac{F2 - HXY1}{\max(HX, HY)};$
Coefficiente de Variação	$F16 = \frac{\sigma(P(i, j))}{\mu(P(i, j))};$
Medida de Probabilidade	$F17 = \max(P(i, j));$
Variância Diagonal	$F18 = \text{var}(P(i, j));$
Momento Diagonal	$F19 = \sum_{i,j} (0.5 i - j P(i, j))^{\frac{1}{2}};$
Segundo Momento Diagonal	$F20 = \sum_{i,j} (0.5 i - j P(i, j));$
Simetria Triangular	$F21 = \sum_{i,j} P(i, j) - P(j, i) ;$

$$\mu_x = \sum_i i \sum_j P(i, j), \quad \mu_y = \sum_j j \sum_i P(i, j);$$

$$\sigma_x = \sum_i (i - \mu_x)^2 \sum_j P(i, j), \quad \sigma_y = \sum_j (j - \mu_y)^2 \sum_i P(i, j);$$

$$P_x(i) = \sum_j P(i, j), \quad P_y(j) = \sum_i P(i, j);$$

$$P_{x+y}(k) = \sum_{i,j|i+j=k} P(i, j), \quad P_{x-y}(k) = \sum_{i,j||i-j|=k} P(i, j);$$

Hx e Hy são entropias de $P_x(i)$ e $P_y(j)$ respectivamente;

$$HXY1 = -\sum_{i,j} P(i, j) \log(P_x(i)P_y(i)).$$

Nesse trabalho serão utilizadas as funções $F1, F2, F3, F4, F5, F6, F7, F8, F9, F10, F13, F14, F15$ e $F21$. Essas funções serão todas calculadas, embora apenas algumas sejam utilizadas de fato, como será mostrado no Capítulo 5, serão escolhidas apenas as mais significativas para a classificação da imagem.

No cálculo de todas as características disponíveis das matrizes de co-ocorrência, para vários valores da distância d e do ângulo de orientação θ , torna-se possível uma análise mais precisa em diferentes resoluções espaciais das imagens, (Conners & Harlow, 1978 e Rosenfeld et al. 1976).

Entretanto, ao se extrair todas as características disponíveis em GLCM, pode resultar em um espaço de grandes dimensões quando comparado ao conjunto de treino disponível para definir ou caracterizar este espaço. Este problema é conhecido como “maldição de dimensionalidade” (*curse of dimensionality*), visto na seção 2.4, e pode ser facilmente mostrado através de um exemplo simples, (Walker, 1997):

Seja \mathbf{X} um vetor bidimensional de variáveis aleatórias normalmente distribuídas, e um vetor \underline{x} bidimensional, onde $\underline{x} = [x_1, x_2]$, representa 100 amostras da variável aleatória \mathbf{X} . O histograma da primeira variável $\underline{x}_1 = [x_{1,1}, \dots, x_{1,100}]^T$, representando a mostra de uma característica de 100 exemplares, é mostrado na Figura 17. Pode ser notado que cada faixa do histograma contém um número de entradas ($100/10 = 10$ medidas por faixa, na média), e a forma do histograma é uma aproximação adequada da verdadeira distribuição de onde as amostras foram extraídas.

A segunda variável $\underline{x}_2 = [x_{2,1}, \dots, x_{2,100}]^T$ representa as amostras da segunda característica dos 100 exemplares. Torna-se necessário estimar este novo histograma bidimensional contendo 100 faixas, com somente o número original (2D) de pontos de dados. Este histograma, mostrado a direita da Figura 17, contém muitas faixas vazias, e todas as outras possuem valores muito baixos. Na média, existem somente $100/100 = 1$ medida por faixa. Como pode ser visto, quanto mais características são extraídas, maior será a imprecisão da estimativa da distribuição multivariável.

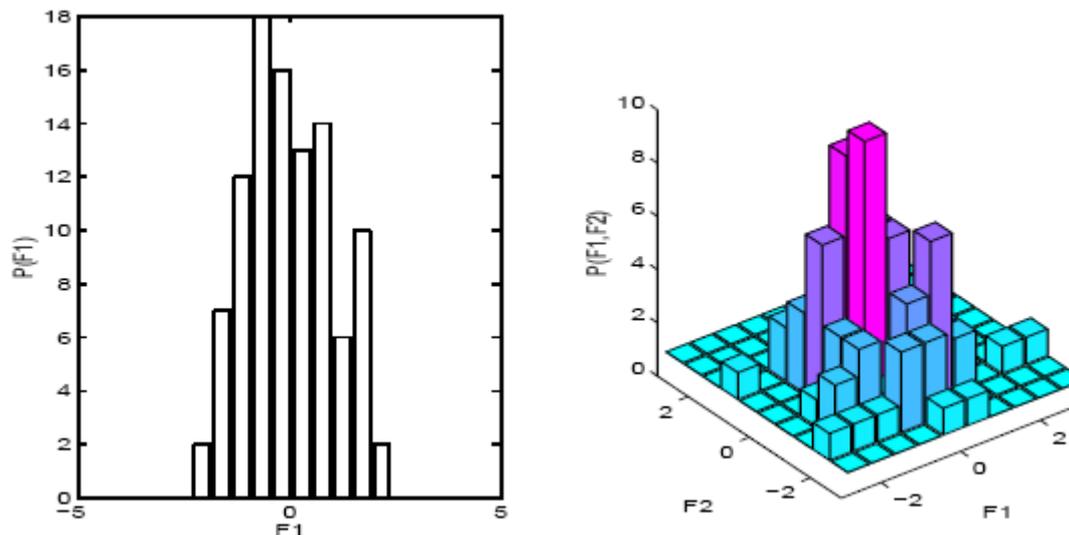


Figura 17 – Um exemplo que mostra como se reduz a exatidão da estimativa da distribuição quando da passagem do espaço de dimensionalidade 1-D para 2-D. Para cada característica, 100 amostras foram desenhadas dos dados distribuídos.

Para fins de classificação, onde texturas desconhecidas deverão ser atribuídas a uma das diversas classes, é de primordial importância a correta estimativa das características de distribuição das classes condicionadas. É desejado se extrair somente um mínimo número de características secundárias “úteis” (onde “úteis” significa características cujas distribuições de suas classes condicionadas possuem diferenças estatísticas entre as mesmas), sendo que, usualmente não é conhecida a princípio quais características são úteis. Uma prática muito usada é a de se extrair um grande número de características secundárias e reduzir o espaço de características, utilizando-se uma análise de discriminantes e técnicas de seleção de características, como vistas na seção 2.6. Isto permite uma melhor estimativa da verdadeira distribuição das características de cada classe, baseada nos dados de treinamento disponíveis.

A Redução da dimensão do conjunto de características, geralmente envolve a remoção daquelas que são redundantes para o processo de classificação. Este processo é efetuado pela remoção das características altamente relacionadas, as quais fornecem muito pouca informação extra e pela remoção das características cujo poder de discriminação é muito baixo.

O ângulo de orientação θ possui significância em alguns casos específicos de matriz de co-ocorrência, como a discriminação de áreas urbanas. Geralmente alguns autores utilizam a média ou a soma dos valores desses parâmetros, que geralmente são calculados para $\theta = 0^\circ$, 45° , 90° , 135° . Nesse trabalho foi utilizada uma técnica para aproveitar somente as componentes dos ângulos que possuem uma certa relevância na classificação das texturas envolvidas, com a utilização das componentes principais – PCA (seção 2.4.1) dos ângulos.

A redução do conjunto de dados se justifica pela quantidade de valores redundantes, presentes em alguns dos 4 ângulos espaciais, calculados para cada função. Como exemplo, pode ser visto na Figura 18, a função de densidade de probabilidade (pdf) do parâmetro $F3$ - Homogeneidade, aplicada a uma imagem de teste. Claramente pode ser observado, que os dados para os ângulos de 45° , 90° e 135° possuem pdf's bastante semelhantes, possuindo assim redundância de informação, que pode ser eliminada pela aplicação de PCA em todos os ângulos espaciais, produzindo assim duas componentes (nesse caso específico – PCA1 e PCA2) fracamente não relacionadas, ver Figura 19.

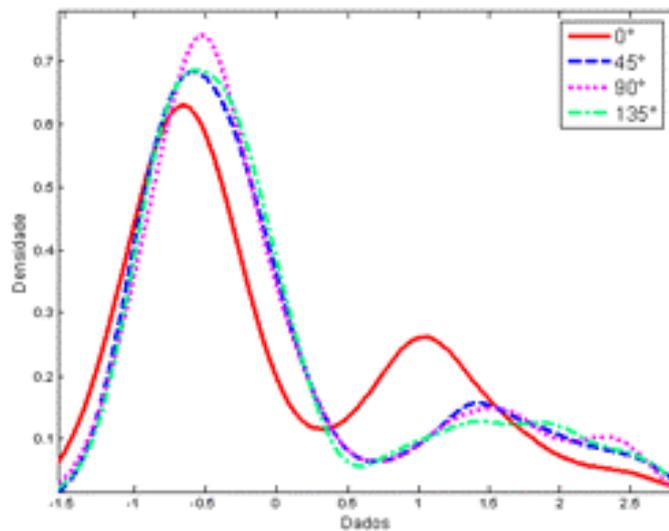


Figura 18 – Função de densidade de probabilidade (pdf) do parâmetro $F3$ - Homogeneidade, com quatro ângulos de orientação θ , aplicada à uma imagem de teste.

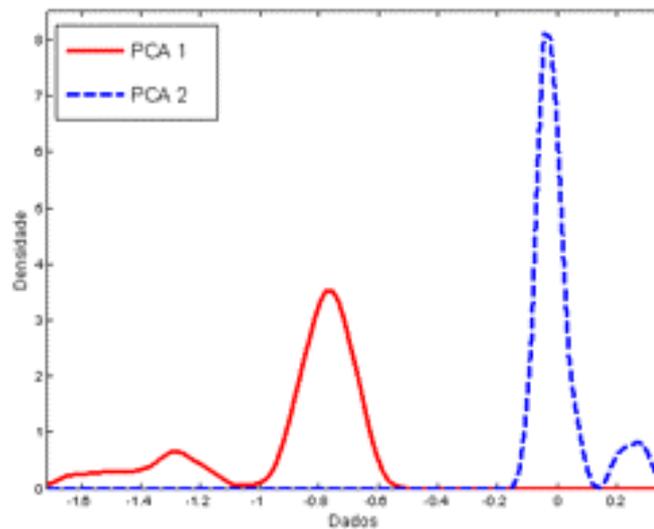


Figura 19 - Aplicação de PCA a pdf da função Homogeneidade – parâmetro $F3$ obtida com quatro ângulos de orientação θ da imagem de teste.

Além disso, o método usado por muitos autores, consiste em se retirar a média ou a soma das características de GLCM desses ângulos o que pode provocar a perda de informações importantes quando da análise total dos dados, podendo uma característica de um ângulo que possibilita uma ótima classificação ser perdida devido à sobreposição de outras que não possuem poder discriminatório. Isso pode ser melhor visualizado por intermédio da Figura 20(a) e (b) que mostram as pdf's dos parâmetros de GLCM espaciais correspondentes a função Entropia ($F2$) com ângulos 0° e 45° aplicadas a uma imagem de teste. Claramente pode ser observado que essa função – $F2$, para esses ângulos possui alto poder discriminatório das classes envolvidas na imagem de teste, característica essa que pode ser perdida caso se faça a soma ou a média desses parâmetros, resultando com isso um baixo poder discriminatório total.

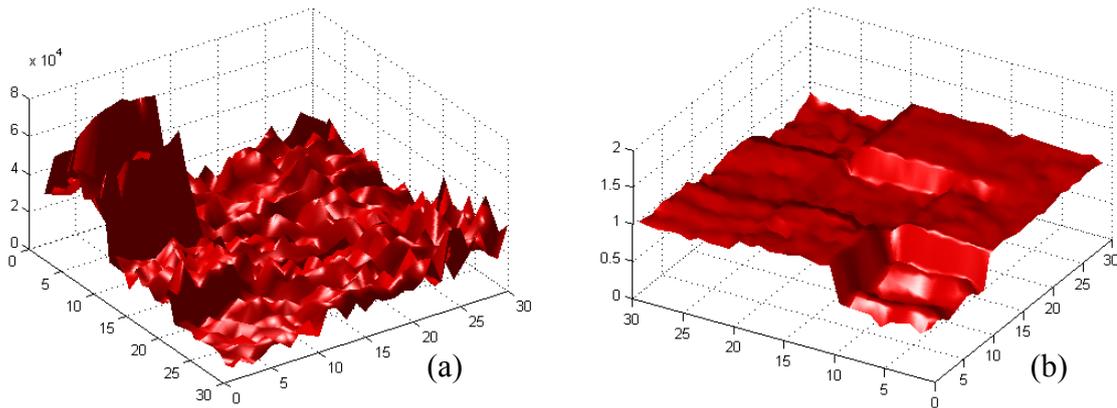


Figura 20 – Função de densidade de probabilidade dos parâmetros espaciais de GLCM correspondentes ao ângulo $\theta = 0^\circ$ (a) e $\theta = 45^\circ$ (b) da função entropia, aplicada a uma imagem de teste.

De outra forma, determinadas funções de GLCM não são muito influenciadas pelo ângulo θ em determinadas texturas, podendo nesses casos ter seus parâmetros somados ou retiradas à média, conforme citado por muitos autores. Ver Figura 21 que mostra a pdf da função energia, para o ângulo $\theta = 0^\circ, 45^\circ, 90^\circ$ e 135° . Todos os ângulos forneceram a mesma pdf para a imagem de teste utilizada. Nesse caso a aplicação da PCA irá resultar em somente uma componente, evitando com isso cálculos computacionais desnecessários e redução do número de entradas da rede neural de classificação.

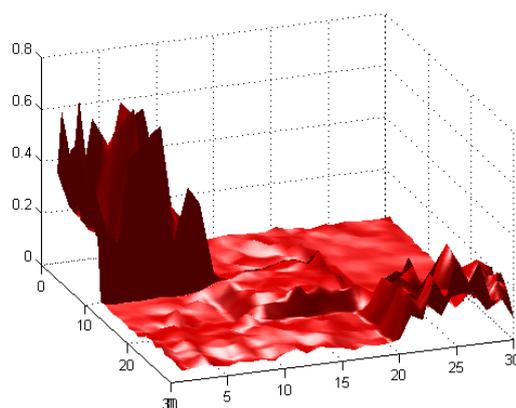


Figura 21 – Função densidade de probabilidade dos parâmetros espaciais de GLCM correspondentes ao ângulo $\theta = 0^\circ, 45^\circ, 90^\circ$ e 135° da função energia, aplicada a uma imagem de teste.

Características de GLCM têm sido extensivamente usadas para classificação de texturas em diversas aplicações. Connors et al. (1984) usaram características de GLCM para segmentação de áreas urbanas em imagens de satélite de alta definição; Schwartz (2003) utilizou para classificação de terrenos; Marceau et al. (1990) para classificação de cobertura da terra em imagens SPOT; Barber & LeDrew (1991) utilizaram para procura de blocos de gelo no mar; Bin Tian et al. (1999) utilizaram para classificação de nuvens e Walker (1979) utilizou GLCM para classificação automática de câncer em células.

Em Haralick et al. (1973), foram utilizadas características extraídas de matriz de co-ocorrência para classificação de terrenos. As características que eles utilizaram foram às relacionadas com as texturas específicas do trabalho desenvolvido. Eles obtiveram uma taxa de acerto entre 80-90% para diferentes tipos de conjuntos de dados. Embora eles tenham usado somente algumas das características, listadas na Tabela 1, e não tenham usado o mesmo tipo de algoritmo para classificação, em todos os conjuntos de dados, pode ser concluído que características extraídas de matrizes de co-ocorrência possuem bom desempenho, em distinguir diferentes tipos de classes de textura, para vários tipos de imagens.

Estudos comparando o desempenho de características extraídas de GLCM com outras técnicas de análise de textura mostraram que GLCM é um dos métodos mais eficientes para classificação geral utilizando texturas, (Connors et al., 1984 e Chen, Nixon & Thomas, 1995). Unser (1986), e Ojala, Pietikainen, & Harwood (1996) concluíram, que outros métodos, como modelos de Campos Aleatórios de Markov, (Chellappa & Khotanzad, 1982), podem ser superiores a GLCM em alguns casos restritos, mas falham quando aplicados a problemas mais gerais envolvendo texturas.

3.3 Distribuição Estatística de Formas Geométricas - SGF

Recentemente, alguns autores como Chen, Nixon & Thomas (1995) tem proposto um novo conjunto de 16 características para classificação com texturas, método este denominado Distribuição Estatística de Formas Geométricas – SGF (“*Statistical Geometric Features*”), cujas propriedades mais significativas são:

- permite uma melhor taxa de classificação que outros métodos usualmente utilizados – como GLCM e Transformada de Fourier (“*Fourier Power Spectral Method*”) Liu & Jernigan 1990;

- a redução de desempenho na classificação com o uso de SGF pelo aumento do número de classes de texturas é menor do que com os outros métodos;
- o desempenho de SGF não se altera com a adição de ruído.

Esse método consiste em decompor, em vários níveis, uma imagem texturizada em escala de cinza em uma seqüência de imagens binárias¹⁶. Algumas propriedades das regiões da imagem (frente e fundo) de cada imagem binária são então mensuradas e um número de parâmetros estatísticos, baseados em propriedades geométricas, são extraídos.

Os métodos de Chen e outros foram baseados na seleção do número de regiões conectadas e irregularidades mensuradas entre duas regiões distintas. Sendo as 16 características mencionadas, obtidas pela multiplicação dessas propriedades por 2 (regiões de frente e fundo da imagem) e logo a seguir pela multiplicação por 4 (parâmetros estatísticos). Existem muitas outras formas possíveis de extração das propriedades estatísticas em imagens binárias, de forma a maximizar o desempenho em um conjunto de texturas em particular.

3.3.1 Algoritmo da Distribuição Estatística de Formas Geométricas

Uma imagem em escala de cinza pode ser modelada em um domínio $D \subset \mathbb{Z}^2$, de N_g níveis de cinza, como uma função em duas dimensões $I: D \rightarrow G$, onde $G = \{1, \dots, N_g\}$. O algoritmo SGF é dado por:

Passo1 – Uma seqüência de imagens binárias $I_b(x, y; \tau)$ é produzida a partir de $I(x, y)$ pelo limiar de cada nível de intensidade discreta $\tau \in \{1, 2, \dots, N_g\}$. Cada imagem binária $I_b(x, y; \tau)$ é obtida por:

$$I_b(x, y; \tau) = \begin{cases} 1 & \text{se } I(x, y) \geq \tau; \\ 0 & \text{caso contrário,} \end{cases} \quad (9)$$

O mapeamento do espaço da imagem discreta em escala de cinza para o espaço binário é uma função bijetiva (“um-para-um” e “sobre”). Onde “um-para-um” significa que cada ponto distinto no espaço da imagem I , é mapeado para uma imagem binária distinta

¹⁶ Imagens constituídas de planos de bits.

$I_b(x, y; \tau)$. O termo “sobre” significa que cada possível ponto no espaço $I_b(x, y; \tau)$ é uma imagem de um ponto em I . Não há perda de informação quando da representação de uma imagem em escala de cinza em uma seqüência de imagens binárias, porque sempre podemos recuperar a imagem original com:

$$I(x, y) = \sum_{\tau=1}^{N_g} I_b(x, y) \rightarrow (x, y) \in D. \quad (10)$$

Para cada imagem binária $I_b(x, y; \tau)$, um grupo de *pixels* de valor 1 é definido como uma região 4-conectada (mais detalhes são mostrados no Apêndice C), se, para todos os *pixels* no grupo, cada *pixel* possui pelo menos um vizinho 4-conectado dentro do grupo. Grupos de *pixels* de valor 0 são definidos de forma similar.

Passo 2 – Uma propriedade geométrica é mensurada para cada região 4-conectada em cada imagem binária. São então tomadas a media ou a soma dessas medidas para todas as regiões de valor 1 e de valor 0 em cada limiar, de forma a obter um par de propriedades geométricas $g_{1(\tau)}, g_{0(\tau)}$ como funções do limiar τ .

Passo 3 – Varias medidas estatísticas que caracterizam as distribuições de $g(\tau)$ para todos os valores de τ , são então computadas. Essas estatísticas são então utilizadas para classificação das texturas.

Foram utilizadas por Chen et al, dois conjuntos de propriedades geométricas. Sendo a primeira uma simples contagem do número de regiões conectadas:

$$NC_1(\tau) = \text{número de regiões 4-conectadas de valor 1} \quad (11)$$

$$NC_0(\tau) = \text{número de regiões 4-conectadas de valor 0} \quad (12)$$

A segunda, uma media do tamanho das regiões, das irregularidades ou não-circularidade das regiões, definida como:

$$\overline{TRGL}_1(\tau) = \frac{\sum_{j=1}^{NC_1(\tau)} IRGL_j(\tau) \cdot NOP_j(\tau)}{\sum_{j=1}^{NC_1(\tau)} NOP_j(\tau)}; \quad (13)$$

$$\overline{TRGL}_0(\tau) = \frac{\sum_{j=1}^{NC_0(\tau)} IRGL_j(\tau) \cdot NOP_j(\tau)}{\sum_{j=1}^{NC_0(\tau)} NOP_j(\tau)}, \quad (14)$$

onde o índice j é a j -ésima região 4-conectada, $NOP_j(\tau)$ é o número de *pixels* na j -ésima região de nível τ , e $IRGL_j(\tau)$ é a irregularidade ou não-circularidade de cada região dada por:

$$IRGL = \frac{1 + \sqrt{\pi} \cdot \frac{\max_{i \in I} \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}}{\sqrt{|R|}}}{\sqrt{|R|}} - 1, \quad (15)$$

onde,

$$\bar{x} = \frac{\sum_{i \in R} x_i}{|R|}, \bar{y} = \frac{\sum_{i \in R} y_i}{|R|}, \quad (16)$$

R é o conjunto de todos os índices de *pixels* na região, e $|R|$ é a cardinalidade, ou número de índices no conjunto R . As características dessas medidas são vistas no Apêndice C.

As quatro funções de nível de limiar τ definidas acima ($NC_1(\tau), NC_0(\tau), \overline{IRGL}_1(\tau), \overline{IRGL}_0(\tau)$) representam distribuições estatísticas. Podendo ser caracterizadas por:

$$\text{valor máximo} = \frac{\max}{\tau} g(\tau), \quad (17)$$

$$\text{valor médio} = \frac{1}{N_g - 1} \sum_{\tau} g(\tau), \quad (18)$$

$$\text{amostragem da média} = \frac{1}{\sum_{\tau} g(\tau)} \sum_{\tau} r \cdot g(\tau), \quad (19)$$

$$\text{amostragem do desvio padrão (D.P.)} = \sqrt{\frac{1}{\sum_{\tau} g(\tau)} \sum_{\tau} (\tau - \text{amostragem da média})^2 \cdot g(\tau)}, \quad (20)$$

onde $g(\tau)$ é uma das quatro funções características. Todas essas funções dão um total de 16 características baseadas nas estatísticas das propriedades geométricas da imagem.

De acordo com diferentes níveis de limiar τ escolhidos e também do nível de quantização da imagem em escala de cinza, ver Figura 22, várias imagens binárias podem ser geradas e utilizadas nas funções descritas acima.

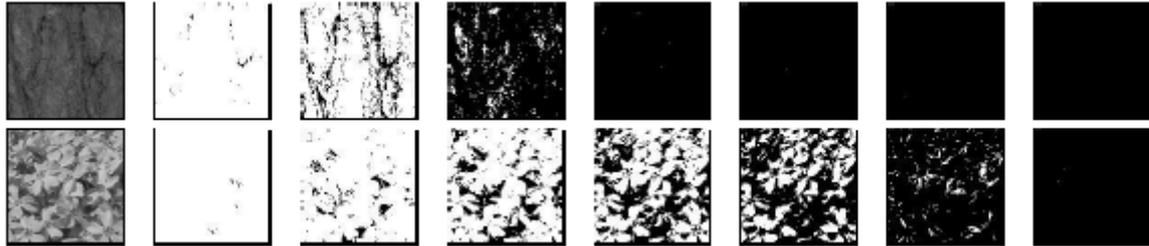


Figura 22 – Exemplo de imagens binárias, produzidas com diferentes níveis de limiar τ aplicadas a duas imagens em 256 níveis de cinza, resultando em 7 imagens binárias.

Capítulo 4

Redes Neurais para Classificação

4.1 Introdução

Um dos ramos da inteligência artificial que mais se desenvolveu ultimamente é o de redes neurais (RNAs). Isso se deve a capacidade de não se precisar fazer suposições mais delicadas a respeito da distribuição dos dados de entrada do que métodos estatísticos tradicionais, ou seja, o seu uso não requer nenhum conhecimento prévio sobre a distribuição estatística dos dados a serem analisados, além de sua capacidade de formar fronteiras de decisão altamente não-lineares no espaço de características. Isso aliado à sua facilidade de utilização levou ao seu crescente uso, especialmente em situações de reconhecimento de imagens.

O desempenho de uma rede neural depende significativamente de quão bem é feito o seu treinamento. Durante a fase de treinamento a rede neural “aprende” a reconhecer padrões presentes no conjunto de treinamento e baseado nesse conjunto constrói regras que podem ser utilizadas para determinar a natureza de dados desconhecidos.

Um neurônio artificial é uma estrutura relativamente simples, que responde a estímulos de outros neurônios conectados a ele, ver Figura 23. Essa arquitetura dá as redes neurais características marcantes de intenso paralelismo e robustez, (Cordeiro, 2002).

Na maioria das aplicações de RP usando-se redes neurais, são estabelecidas conexões entre os valores de todas as características que definem os padrões com camadas intermediárias de neurônios e a todas as classes objetivos que são consideradas como a camada de saída.

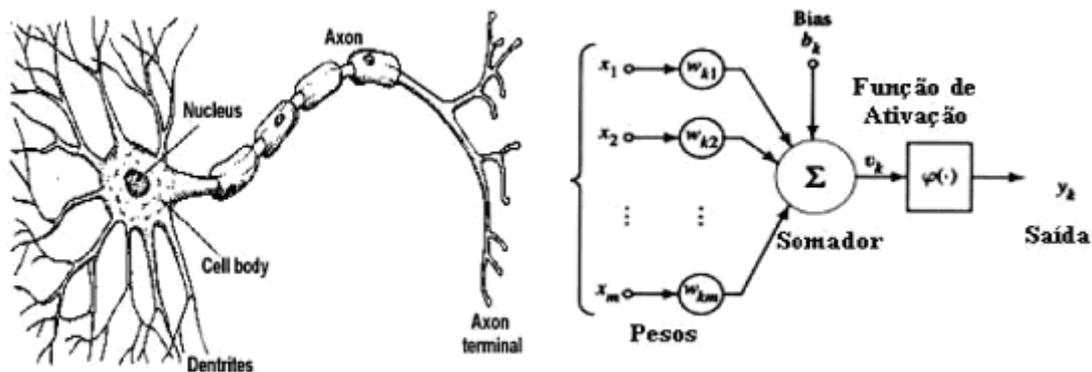


Figura 23 - Representação artificial de um neurônio natural. As entradas x_1, x_2, \dots, x_n , representam os sinais vindos de outros n neurônios. Os pesos $w_{k1}, w_{k2}, \dots, w_{kn}$ representam os pesos sinápticos das conexões entre os neurônios da camada anterior e da camada k . Uma entrada de polarização (*bias*) é o limiar para a ativação do neurônio.

O treinamento da rede é realizado corrigindo os pesos nas conexões para estabelecer-se as relações entre as características e classes que promovam a melhor discriminação possível entre os padrões de classes diferentes. Assim, ao ser apresentado à rede um novo padrão, esta indicará a classe que o representa melhor na camada de saída, (Artur, 1999).

Grande parte do trabalho de quem faz uso de uma rede neural está em determinar a arquitetura a ser utilizada na rede, e a definição dos parâmetros inerentes as redes, como a taxa de aprendizado, que influencia no seu tempo de treinamento. Não existem regras claras de forma a construir uma rede neural, somente heurísticas de forma a guiar os usuários na escolha do tipo e dos parâmetros das redes.

Vários tipos de arquiteturas de redes neurais tem sido utilizadas para a classificação de imagens, incluindo as redes multicamadas (MLP – *Multi-Layer Perceptron*) com algoritmo de aprendizagem *backpropagation*, a rede LVQ (Learning Vector Quantization), os mapas topológicos de Kohonen (SOM - *Self-Organizing Map*), e mais recentemente a rede de função de base radial (RBF – *Radial Basis Function*) e sua variação a rede probabilística (PNN – *Probabilistic Neural Network*).

4.2 Rede Neural Multicamadas

As redes MLPs têm sido aplicadas com sucesso em uma variedade de áreas, desempenhando tarefas tais como: classificação de padrões (reconhecimento), controle e processamento de sinais. Uma RNA do tipo MLP é constituída por um conjunto de nós fonte, os quais formam a camada de entrada da rede (*input layer*), uma ou mais camadas escondidas (*hidden layers*) e uma camada de saída (*output layer*). Com exceção da camada de entrada, todas as outras camadas são constituídas por neurônios e, portanto, apresentam capacidade computacional.

A Figura 24 mostra a arquitetura de uma rede neural MLP com uma camada de entrada, duas camadas escondidas e uma camada de saída.

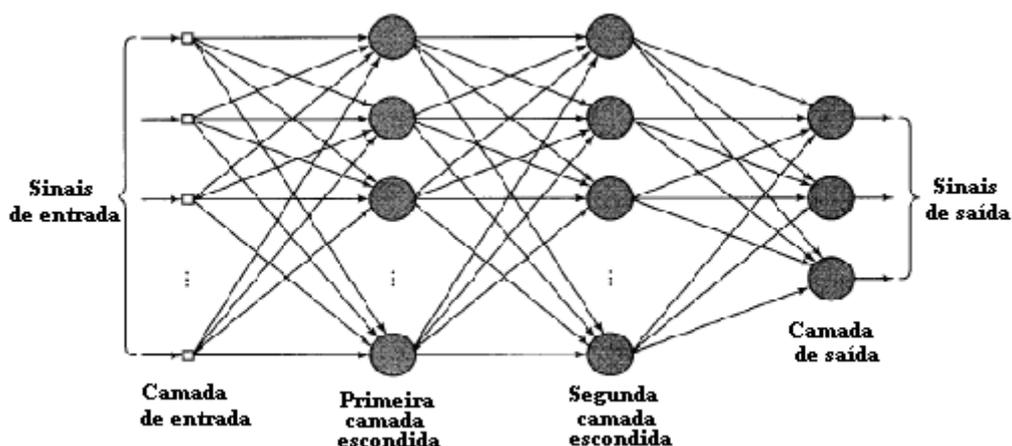


Figura 24 – Arquitetura de uma rede neural MLP com duas camadas escondidas.

Uma rede MLP é uma rede progressiva (*feedforward*), ou seja, as saídas dos neurônios em qualquer camada particular se conectam unicamente às entradas dos neurônios da camada seguinte, sem a presença de laços de realimentação. Conseqüentemente, o sinal de entrada se propaga através da rede, camada a camada, em um sentido progressivo.

A rede pode ser completamente conectada, caso em que cada nó (computacional ou não) em uma camada é conectado a todos os outros nós da camada adjacente. De forma alternativa, uma rede MLP pode ser parcialmente conectada, caso em que algumas sinapses poderão estar faltando. Neste estudo, no entanto, consideraremos apenas MLPs completamente conectados.

O número de nós fonte na camada de entrada da rede é determinado pela dimensionalidade do espaço de observação, que é responsável pela geração dos sinais de entrada. O número de neurônios na camada de saída é determinado pela dimensionalidade requerida da resposta desejada. Assim, o projeto de uma rede MLP requer a consideração de três aspectos:

1. A determinação do número de camadas escondidas;
2. A determinação do número de neurônios em cada uma das camadas escondidas;
3. A especificação dos pesos sinápticos que interconectam os neurônios nas diferentes camadas da rede.

Os aspectos 1 e 2 determinam a complexidade do modelo de RNA escolhido e, infelizmente, não há regras determinadas para tal especificação. A função das camadas escondidas em uma RNA é a de influir na relação entrada-saída da rede de uma forma ampla. Uma RNA com uma ou mais camadas escondidas é apta a extrair as estatísticas de ordem superior de algum desconhecido processo aleatório subjacente, responsável pelo "comportamento" dos dados de entrada, processo sobre o qual a rede está tentando adquirir conhecimento. A RNA adquire uma perspectiva global do processo aleatório, apesar de sua conectividade local, em virtude do conjunto adicional de pesos sinápticos e da dimensão adicional de interações neurais proporcionada pelas camadas escondidas.

O aspecto 3 envolve a utilização de algoritmos de treino supervisionados. As RNAs MLPs têm sido aplicadas na solução de diversos e difíceis problemas através da utilização de tais algoritmos. O algoritmo de treino quase universalmente utilizado para tanto é o algoritmo de retro-propagação do erro, conhecido na literatura como backpropagation.

Uma rede MLP apresenta três características distintas, de cuja combinação com a habilidade de aprender através da experiência (através do treinamento), deriva sua capacidade computacional:

1. O modelo de cada neurônio do MLP inclui uma função de ativação não-linear. Uma forma comumente utilizada de não-linearidade que satisfaz este requisito é a não-linearidade sigmoidal definida pela função logística:

$$y_j = \frac{1}{1 + \exp(-v_j)}, \quad (21)$$

onde v_j é o potencial de ativação (isto é, a soma ponderada de todas as entradas sinápticas mais a polarização) do neurônio j , e y_j é a saída do neurônio.

2. O MLP contém uma ou mais camadas de neurônios escondidos que não são parte da camada de entrada ou da camada de saída da rede. Estes neurônios escondidos possibilitam que a rede aprenda tarefas complexas, extraindo progressivamente mais características significativas dos padrões de entrada (vetores de entrada).
3. A rede MLP exhibe um alto grau de conectividade, determinado pelas sinapses da rede. Uma mudança na conectividade da rede requer uma mudança na população de conexões sinápticas, ou pesos sinápticos.

Estas mesmas características, entretanto, são também responsáveis pelas dificuldades encontradas na análise de tais redes. Por exemplo, a presença das não-linearidades distribuídas e a alta conectividade tornam difícil a análise teórica das redes MLPs. Em uma rede MLP, o conhecimento aprendido sobre o ambiente é representado pelos valores assumidos pelos pesos sinápticos da rede. A natureza distribuída deste conhecimento ao longo da rede a torna de difícil interpretação. Além disso, o uso de neurônios escondidos torna o processo de aprendizado mais difícil de ser "visualizado" na estrutura da rede. Mais informações podem ser obtidas em Sing-Tze-Bow (2002).

4.3 Rede Neural de Função de Base Radial

A determinação de uma função de base radial usada como função de transferência, aplicada aos neurônios escondidos de uma RBF, é o fator primordialmente inovador desta arquitetura de RNA, ilustrada na Figura 25. A utilização de uma RBF favorece o processo de classificação, uma vez que nenhuma forma de distribuição de probabilidade é assumida para o conjunto de dados e a designação de um *pixel* a uma classe é realizada através da consideração, simultânea, dos efeitos de todas as bases radiais que identificam as classes sob estudo.

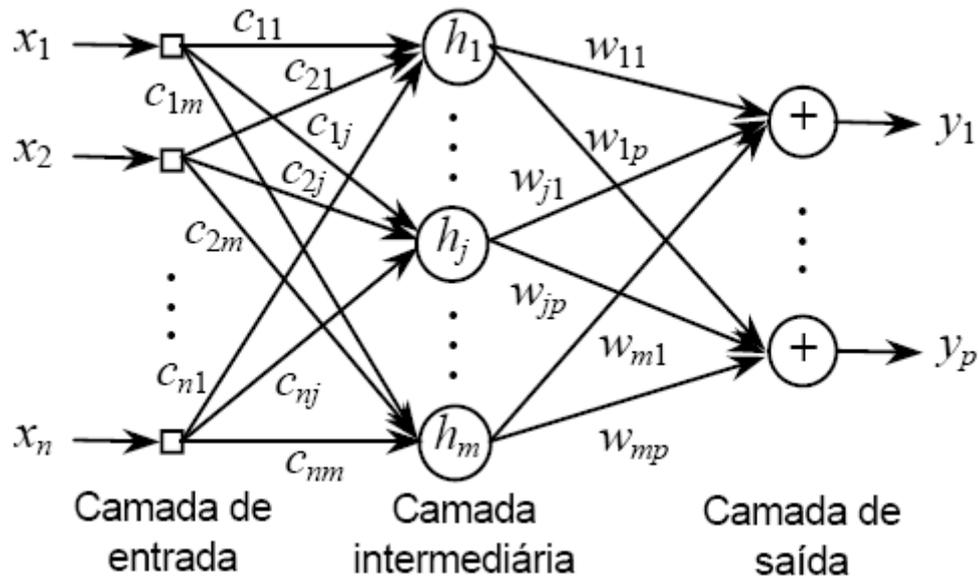


Figura 25 – Arquitetura básica de uma RBF

- a primeira camada é constituída por n entradas ($x_i, i = 1, \dots, n$);
- a segunda, por M neurônios escondidos associados a M centros, $h_j (j = 1, \dots, m)$, e
- a terceira, uma camada constituída por p neurônios de saída $y_k (k=1, \dots, p)$.

As funções de base radial (funções de transferência da segunda camada) aplicadas aos M centros, h_j , são normalmente, funções Gaussianas, cujos resultados alimentam a camada de p neurônios de saída. Segundo Bishop (1995), a função Gaussiana, não normalizada, calculada para cada uma das n entradas, pode ser obtida por:

$$\varphi(x_{jn}) = \exp\left(-\frac{\|x_{jn} - h_j\|^2}{2\sigma_j^2}\right), \quad (22)$$

onde x_{jn} é o n -ésimo vetor de entrada p -dimensional que alimenta o neurônio j , com $n=1, \dots, N$; h_j é o vetor correspondente ao j -ésimo centro p -dimensional, definido no período de treinamento, com $j=1, 2, \dots, M$ e σ_j^2 é a variância presente neste j -ésimo centro (o desvio padrão define o campo receptivo de cada centro; Gaussiana só apresentará resultados significativos para entradas inclusas neste campo receptivo).

A camada de saída em uma RBF tem como principal função reduzir a dimensionalidade de classes existentes na camada escondida. A partir dessa característica, a RBF possibilita que *pixels* representando uma determinada classe, mas que apresentem variações nos níveis de cinza, possam ser classificadas corretamente. A função de transferência normalmente

utilizada na camada de saída é a função identidade. A utilização dessa função é muito importante uma vez que permite obtermos os pesos de forma direta, Bishop (1995).

O treinamento de uma RBF é híbrido e ocorre em duas etapas. Na primeira etapa, os pesos são treinados de forma não supervisionada e na segunda, os pesos são obtidos através de um processo supervisionado. De forma geral, o aprendizado dos pesos na primeira etapa exige grande esforço computacional devido ao elevado número de operações requeridas.

4.3.1 Rede Neural Probabilística - PNN

Dados provenientes de imagens naturais, geralmente não possuem distribuição estatística normal, ocasionando com isso dificuldades na determinação de sua distribuição característica. O sucesso de qualquer análise recai sobre a habilidade em se aproximar sua distribuição real. Distribuições gaussianas multivariáveis podem apresentar falhas ao capturar os detalhes de uma classe que possui um número de agrupamentos distintos no seu espaço amostral. Nesse caso, um Modelo de Mistura Gaussiana¹⁷ (*Gaussian Mixture Model - GMM*) pode ser uma escolha apropriada para uma correta classificação, (Raghu e Yegnanarayana, 1998).

A determinação de um modelo de mistura consiste em estimar os parâmetros e os pesos de cada componente. Um algoritmo de maximização chamado “*expectation-maximization (EM)*” tem sido muito utilizado no cálculo computacional da estimativa em se estimar a probabilidade máxima dos parâmetros, levando-se em consideração os dados de forma incompleta, (Dempster, 1977). Esta aproximação é geralmente muito complexa e requer um grande cálculo computacional. Alternativamente, métodos baseados em redes neurais artificiais são mais apropriados para o cálculo do modelo de misturas dos parâmetros envolvidos, sendo possível capturar este modelo utilizando-se uma rede PNN, (Donald, 1990).

A rede neural probabilística é uma extensão da rede RBF, quando uma entrada é apresentada, a primeira camada computa a distância do vetor de entrada aos vetores de treinamento, produzindo um vetor cujos elementos indicam a distância do vetor de entrada para os vetores de treinamento (classes). A segunda camada soma a contribuição da entrada para cada classe, produzindo em sua saída um vetor de probabilidades. Finalmente, uma função de transferência competitiva encontra a classe com maior probabilidade de representar o vetor de entrada, mais detalhes podem ser vistos no Apêndice B.

¹⁷ Modelo cujas variáveis independentes são frações de um total.

Um classificador PNN possui uma relativa facilidade de implementação e treinamento, tornando-o ideal para um sistema automatizado de classificação.

As redes neurais artificiais do tipo PNN são estruturas de rede que, usando campos receptores locais, executam funções de mapeamento. Treinar este tipo de rede corresponde a encontrar uma superfície em um espaço multidimensional que provê aos dados de treinamento o melhor ajuste do ponto de vista estatístico.

Trabalhos realizados com esse tipo de rede incluem os de Nishida e Bastos (1998), Raghu e Yegnanarayana (1998) e Yiu et al. (1999).

4.4 Metodologia de Construção das Redes Neurais Utilizadas

Segundo Tatibana & Kaetsu, deve-se utilizar alguns procedimentos, ou passos, na construção das rede neurais, permitindo com isso um melhor aproveitamento das técnicas utilizadas. Esses procedimentos são descritos da seguinte forma:

4.4.1 Coleta de Dados e Separação em Conjuntos

Os dois primeiros passos do processo de desenvolvimento de redes neurais artificiais são a coleta de dados relativos ao problema e a sua separação em um conjunto de treinamento e um conjunto de testes. Esta tarefa requer uma análise cuidadosa sobre o problema para minimizar ambigüidades e erros nos dados. Além disso, os dados coletados devem ser significativos e cobrir amplamente o domínio do problema; não devem cobrir apenas as operações normais ou rotineiras, mas também as exceções e as condições nos limites do domínio do problema.

Normalmente, os dados coletados são separados em duas categorias: dados de treinamento, que serão utilizados para o treinamento da rede e dados de teste, que serão utilizados para verificar o desempenho sob condições reais de utilização. Além dessa divisão, pode-se usar também uma subdivisão do conjunto de treinamento, criando um conjunto de validação, utilizado para verificar a eficiência da rede quanto a sua capacidade de generalização durante o treinamento, e podendo ser empregado como critério de parada do treinamento.

Depois de determinados estes conjuntos, eles são geralmente colocados em ordem aleatória para prevenção de tendências associadas à ordem de apresentação dos dados. Além disso, pode ser necessário pré-processar estes dados, através de normalizações, escalonamentos e conversões de formato para torná-los mais apropriados à sua utilização na rede.

4.4.2 Configuração da rede

O terceiro passo é a definição da configuração da rede, que pode ser dividido em três etapas:

1. Seleção do paradigma neural apropriado à aplicação.
2. Determinação da topologia da rede a ser utilizada - o número de camadas, o número de unidades em cada camada, etc.
3. Determinação de parâmetros do algoritmo de treinamento e funções de ativação. Este passo tem um grande impacto no desempenho do sistema resultante.

Existem metodologias, "dicas" e "truques" na condução destas tarefas. Normalmente estas escolhas são feitas de forma empírica. A definição da configuração de redes neurais é ainda considerada uma arte, que requer grande experiência dos projetistas.

4.4.3 Treinamento

O quarto passo é o treinamento da rede. Nesta fase, seguindo o algoritmo de treinamento escolhido, serão ajustados os pesos das conexões. É importante considerar, nesta fase, alguns aspectos tais como a inicialização da rede, o modo de treinamento e o tempo de treinamento.

Uma boa escolha dos valores iniciais dos pesos da rede pode diminuir o tempo necessário para o treinamento. Normalmente, os valores iniciais dos pesos da rede são números aleatórios uniformemente distribuídos, em um intervalo definido. A escolha errada destes pesos pode levar a uma saturação prematura.

Quanto ao modo de treinamento, na prática é mais utilizado o modo padrão devido ao menor armazenamento de dados, além de ser menos suscetível ao problema de mínimos locais, devido à pesquisa de natureza estocástica que realiza. Por outro lado, no modo *batch* se tem uma melhor estimativa do vetor gradiente, o que torna o treinamento mais estável.

Quanto ao tempo de treinamento, vários fatores podem influenciar a sua duração, porém sempre será necessário utilizar algum critério de parada. O critério de parada do algoritmo backpropagation não é bem definido, e geralmente é utilizado um número máximo de ciclos. Mas, devem ser considerados a taxa de erro médio por ciclo, e a capacidade de generalização da rede. Pode ocorrer que em um determinado instante do treinamento a generalização comece a degenerar, causando o problema de over-training, ou seja, a rede se especializa no conjunto de dados do treinamento e perde a capacidade de generalização.

O treinamento deve ser interrompido quando a rede apresentar uma boa capacidade de generalização e quando a taxa de erro for suficientemente pequena, ou seja, menor que um erro admissível. Assim, deve-se encontrar um ponto ótimo de parada com erro mínimo e capacidade de generalização máxima.

4.4.4 Teste

O quinto passo é o teste da rede. Durante esta fase o conjunto de teste é utilizado para determinar o desempenho da rede com dados que não foram previamente utilizados. O desempenho da rede, medida nesta fase, é uma boa indicação de seu desempenho real.

Capítulo 5

Testes e Experimentos

5.1 Introdução

Este capítulo apresenta alguns resultados que mostram o desempenho das técnicas propostas. Foram feitos experimentos com imagens reais e sintéticas, a saber: imagens de texturas extraídas da biblioteca Mit's VisTex (ver Apêndice D) e imagem de satélite.

A Figura 26 mostra em (a) três imagens de teste (P1c, P2c, P3c) contendo diferentes tipos de texturas, baseadas na biblioteca VisTex, enquanto em (b) podem ser vistas essas mesmas imagens, agora denominadas P1, P2 e P3, em escala de cinza, obtidas pela componente V (*Value* – valor) usando uma transformação do espaço de cores RGB para o HSV (ver seção 2.3.2.1). A imagem de sensoriamento remoto, representando uma parte da cidade do Rio de Janeiro, é mostrada na Figura 27, denominada de S1, obtida pelo satélite Landsat 7, bandas espectrais 3, 4 e 5, registradas respectivamente pelos canais 3, 4 e 5 do sensor TM. A imagem recebeu um tratamento de contraste conforme visto na seção 2.2, para melhor visualização. A imagem mostrada na Figura 27b, chamada de S2, representa a mesma imagem de satélite em escala de cinza (componente V do espaço de cores HSV), após a aplicação da mesma transformação no espaço de cores utilizada nas imagens sintéticas. Para efeito de classificação foram consideradas três classes nessa imagem: água, área construída e vegetação.

A Tabela 2 ilustra a quantidade de classes diferentes para cada imagem de teste.

O foco desse capítulo se concentra na comparação entre os métodos de extração de características de texturas com GLCM e SGF, como visto no Capítulo 3, embora os experimentos utilizados nesse capítulo possuam outras vertentes. São investigados também: na seção 5.2, o tipo de classificador que apresenta o melhor resultado, efetuando uma comparação entre uma rede neural multicamadas - MLP e uma rede probabilística – PNN.

Na comparação dos diversos experimentos realizados foram utilizados os coeficientes de Kappa obtidos das matrizes de erro das imagens classificadas.

Na seção 5.3 foram feitos vários testes para avaliação do método GLCM para extração de características da imagem, com ênfase na análise dos diversos tipos de funções características utilizadas em GLCM, seção 3.2.1.

A seção 5.4 avalia o método SGF, na extração de características, nas mesmas imagens utilizadas na seção 5.3, para fins de comparação posterior.

A seção 5.5 faz uma comparação entre os métodos de extração de características GLCM e SGF expostos nas seções anteriores.

Finalmente na seção 5.6 são feitos diversos testes com o uso de informações de cores das imagens.

Todos os procedimentos necessários para realizar os testes e experimentos para avaliar as metodologias propostas foram programados utilizando o software Matlab.

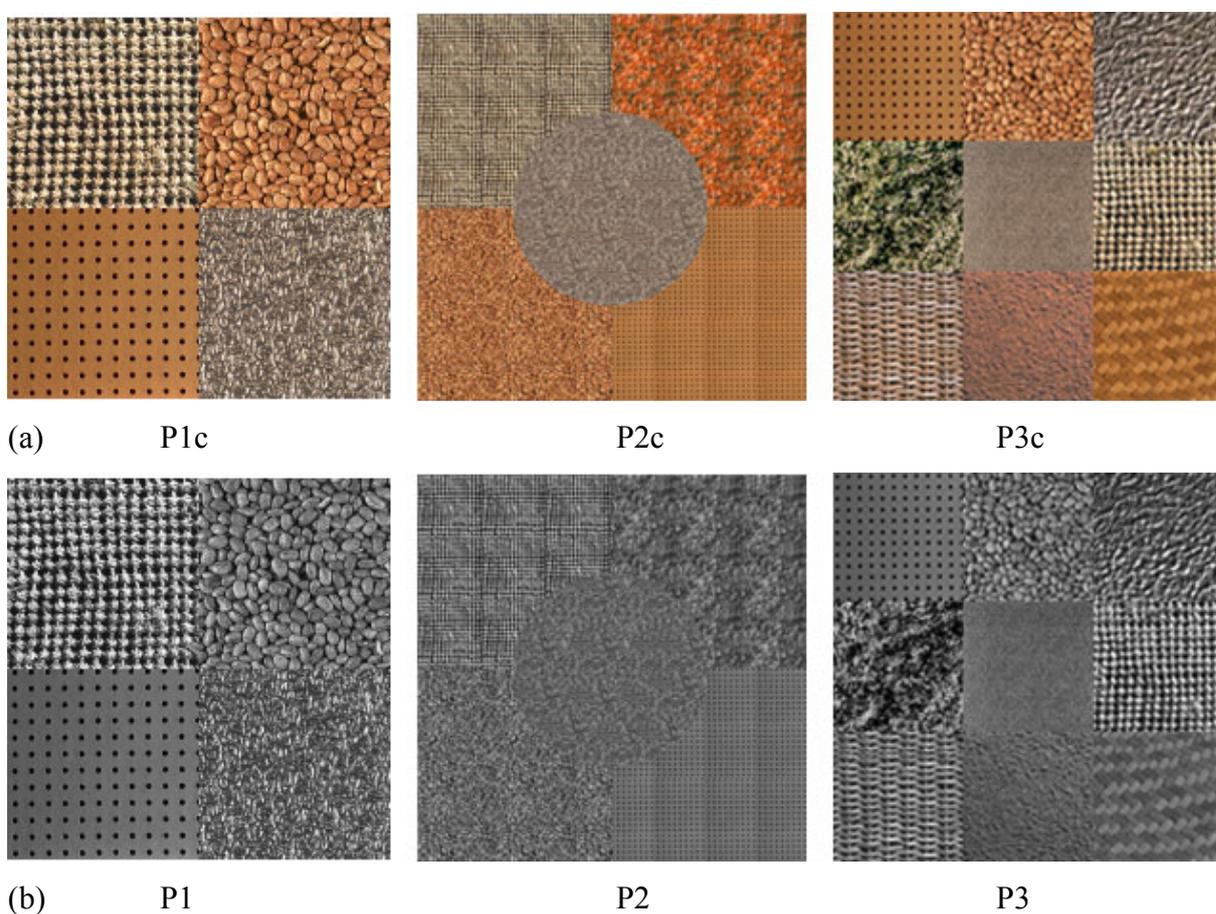


Figura 26 – Imagens de teste construídas com a biblioteca VisTex, formadas pelas texturas: Fabric.0009, Fabric.0014, Fabric.0016, Food.0000, Food.0005, Leaves.0003, Metal.0000, Metal.0002, Metal.0004, Tile.0009. O conjunto (a) mostra as texturas coloridas enquanto que no conjunto (b) são mostradas em escala de cinza. As imagens P1c, P1, P2c, P2 tem tamanho 200x200 *pixels* enquanto as imagens P3c e P3 possuem 500x500 *pixels*.

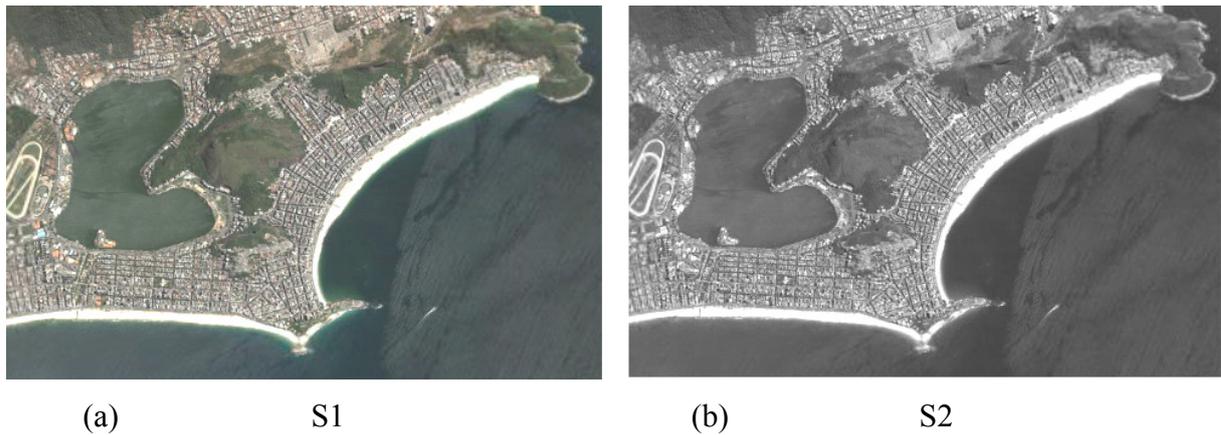


Figura 27 – Imagem real de teste, representando uma parte da cidade do Rio de Janeiro, tamanho 490 x 309 *pixels*, em (a) imagem original colorida e em (b) escala de cinza.

Fonte: Satélite Landsat 7, bandas espectrais 3, 4 e 5, registradas respectivamente pelos canais 3, 4 e 5 do sensor TM.

Tabela 2 - Número de classes distintas das imagens de teste.

Imagem	Nº classes
P1, P1c	4
P2, P2c	5
P3, P3c	9
S1, S2	3

5.2 Escolha do Classificador

É extenso o uso das redes MLP na abordagem de classificação supervisionada, porém recentemente outros trabalhos tem sido feitos com o uso das redes de base radial como a rede neural de função probabilística – PNN.

Nesta seção é analisado o desempenho de uma rede PNN em comparação com a rede MLP, com os dados obtidos das imagens de teste ilustradas na Figura 26b. Será demonstrado que uma rede PNN produz um melhor resultado na classificação das imagens em análise.

As redes PNN's utilizadas, uma para cada imagem de teste, possuem os mesmos números de entrada que as redes MLP's utilizadas, uma camada escondida e uma quantidade

de neurônios de saída de acordo com o número de classes de cada imagem de teste, ver seção 4.2.1. Foram utilizados os mesmos conjuntos de treinamento para a rede PNN, sendo o algoritmo de treino de uma rede PNN bastante trivial (passo único), em comparação com o de uma rede MLP, isso torna o seu treinamento muito mais rápido.

Os testes foram realizados utilizando os dois métodos de extração de características, GLCM e SGF.

O teste com GLCM, foi realizado com a utilização de seis parâmetros de textura, a saber: energia, inércia, correlação, entropia, variância e homogeneidade (ver Tabela 1) e esquema de quantização uniforme segundo a equação (2) em 16 níveis (N_g). O valor do parâmetro d (distância da GLCM) foi fixado em 1 e os ângulos de orientação, θ , foram escolhidos como 0° , 45° , 90° e 135° , tomando-se o valor médio dos valores dos ângulos escolhidos, de acordo com a equação (6).

A utilização de SGF seguiu os mesmos parâmetros da seção 5.4 com 16 funções extraídas.

As redes neurais de duas camadas – MLP foram treinadas com o algoritmo de aprendizado *backpropagation*. O número m de entradas utilizadas foi fixado em seis para GLCM e em 16 para SGF, de acordo com os parâmetros de cada método utilizado, enquanto que o número de saídas corresponde ao número de classes da imagem a ser classificada. Da mesma forma a rede neural PNN possui os mesmos parâmetros de entrada e saída.

Foram utilizados conjuntos de treinamento, teste e validação obtidos de arquivos de texturas que foram utilizados na construção das imagens originais, ver Figura 26b, extraíndo-se desses arquivos parâmetros estatísticos de textura, com janelas em posições aleatórias, conforme ilustrado na Figura 28. Essa abordagem de extração do conjunto de treino é útil na construção de um banco de texturas para segmentação de imagens.

Conforme descrito na seção 4.4, todos os dados extraídos foram normalizados¹⁸ antes de serem aplicados às redes neurais.

A Figura 29 mostra o esquema utilizado para os testes. Após a extração de características de textura das imagens de teste, de acordo com o método (GLCM e SGF), foram utilizadas para treinamento e classificação com as redes em análise (MLP e PNN) sendo os resultados da classificação utilizados para avaliar as redes (Coeficiente de Kappa).

¹⁸ Dados subtraídos da média do conjunto e divididos pelo desvio padrão.

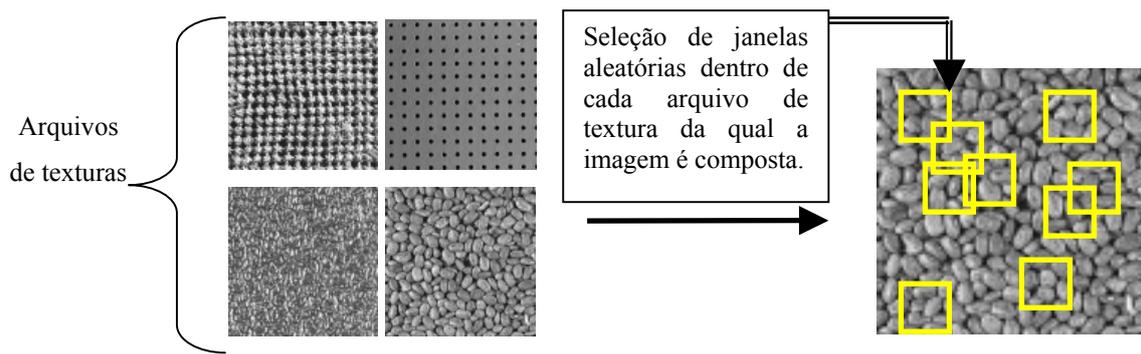


Figura 28 – Seleção do conjunto de treino para aprendizado da rede neural utilizada para classificação das imagens de teste.

Na interpretação da precisão nas classificações, foi utilizado o coeficiente de Kappa (Apêndice A), sendo coeficiente de concordância de Kappa uma medida de precisão para classificação temática que varia de 0 até 1.

Na Tabela 3 e na Figura 30 são mostradas uma comparação, com o uso do coeficiente Kappa, da classificação das imagens de teste, com a utilização das redes neurais MLP e PNN.

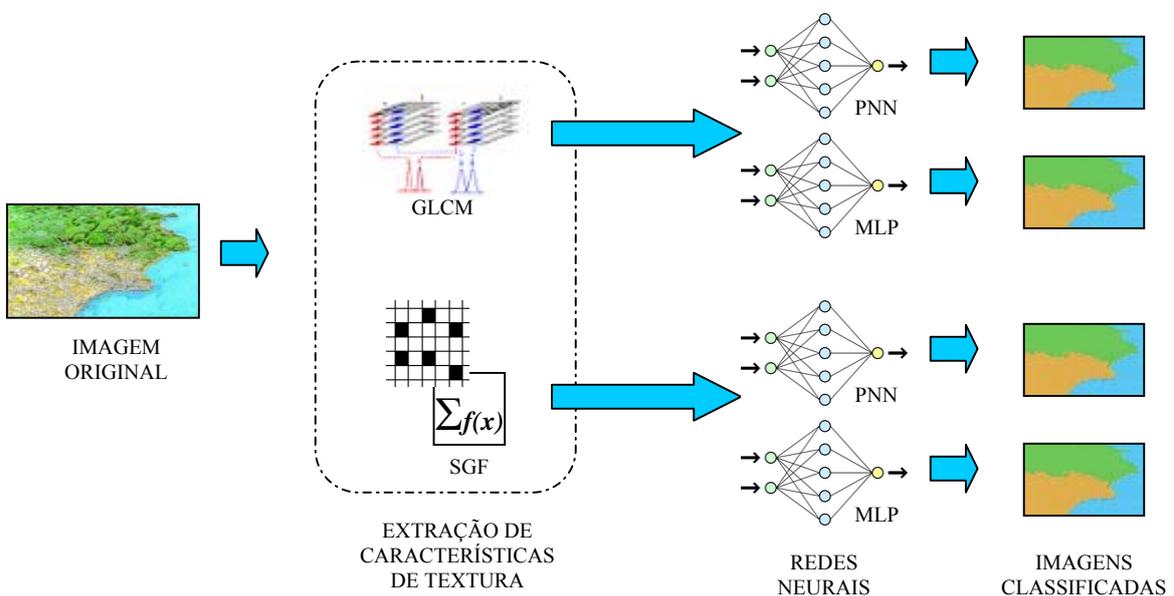


Figura 29 – Esquema utilizado para os testes de escolha do tipo de rede neural.

Tabela 3 – Comparação dos classificadores neurais

Imagem	Kappa			
	GLCM		SGF	
	MLP	PNN	MLP	PNN
P1	0,90	0,92	0,95	0,95
P2	0,81	0,83	0,86	0,88
P3	0,85	0,88	0,89	0,91

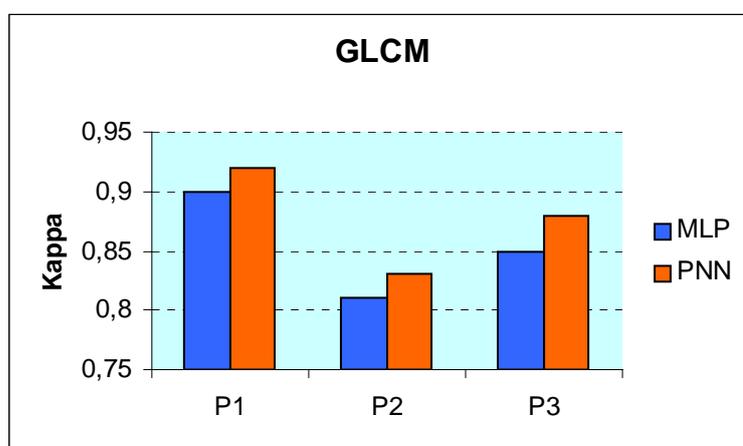


Figura 30 – Comparação de desempenho dos classificadores neurais (MLP e PNN) nas imagens de teste, com o uso de GLCM.

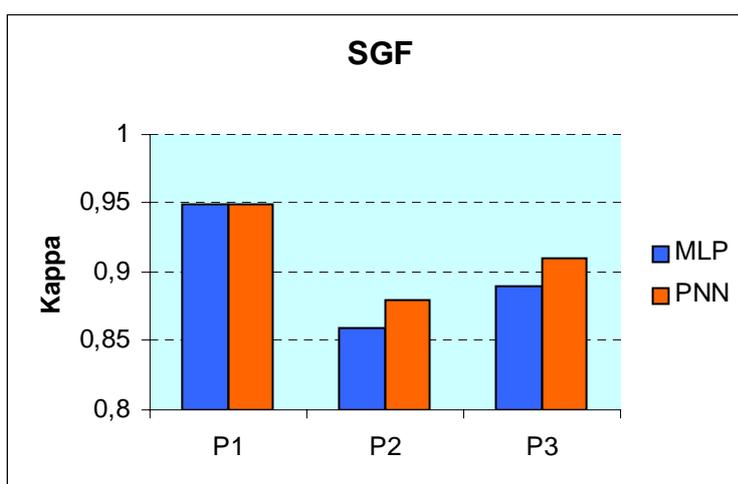


Figura 31 - Comparação de desempenho dos classificadores neurais (MLP e PNN) nas imagens de teste, com o uso de SGF.

Devido ao melhor desempenho, tanto no nível de classificação, índice Kappa, quanto no tempo de treinamento das redes (a rede PNN utiliza um algoritmo de passo único, ao contrário da rede MPL que utiliza o algoritmo, de vários passos, backpropagation), será utilizada somente a rede PNN para os demais experimentos. Deve ser ressaltado que o tempo de treinamento das redes PNN foram bem inferiores ao das redes MLP utilizadas.

5.3 Avaliação e Testes com GLCM

Existem vários parâmetros importantes a considerar quando da utilização de GLCM, como por exemplo, o tamanho da região a ser analisada – janela de amostragem, o nível de quantização de cinza da imagem (N_g – ver seção 2.2), o valor da distância d e do ângulo θ da matriz de co-ocorrência. Outro fator importante se refere ao uso de cores na imagem.

O tamanho da janela de amostragem é um fator crítico, janelas grandes são necessárias para capturar corretamente as características das texturas sob análise, de forma a aumentar a discriminação entre as classes. Entretanto quanto maior a janela, maior poderá ocorrer erros de sobreposição entre duas ou mais regiões no processo de classificação. A isso se chama “efeito de fronteira” (*edge effect*). Para se evitar esse efeito, deve-se reduzir ao máximo o tamanho da janela até o limite da correta separação entre classes, o que geralmente ocorre para diferentes tipos de texturas. Outro problema envolvendo a correta seleção do tamanho da janela, se refere aos diversos tipos de texturas em análise, devendo ser então ser estimada para o grupo de texturas em análise. No presente trabalho o tamanho da janela foi estimado em 10×10 pixels, tamanho este calculado de forma empírica, o qual forneceram os melhores resultados de classificação.

O parâmetro d da GLCM não é tão importante em muitas aplicações. Aplicando um valor alto para esse parâmetro em texturas finas, resulta em valores da matriz de co-ocorrência que não capturam a correta informação da textura e vice-versa. Chen et al. (1992) utilizaram $d = 1, 2, 4, 8, 16, 32, 64$ e mostrou que classificação com os valores $d = 1, 2, 4, 8$ são praticamente equivalentes. Entretanto para valores altos a precisão da classificação decresceu. Ele também conclui que os melhores resultados foram obtidos com valores de $d = 1, 2$. Nesse trabalho foi utilizado o valor de $d = 1$, devido aos tipos de texturas utilizadas, que são texturas finas, além disso, foi dada ênfase ao número de funções diferentes de GLCM (Tabela 1). Resultados semelhantes foram apresentados por Connors et al. (1984), onde raramente são utilizados valores maiores do que $d = 1$.

O número de níveis de cinza é um dos fatores que pode influenciar no cálculo da GLCM. A decisão de quantos níveis utilizar é uma decisão baseada nos tipos de texturas em análise. Quanto maior o nível de quantização (maior valor de N_g) mais informações a respeito das texturas serão extraídas, por outro lado haverá um custo significativo da computação envolvida nesse cálculo. No presente trabalho a quantização uniforme em 16 níveis de cinza ($N_g = 16$), segundo a equação (2), ver seção 2.2, foi a que apresentou melhores resultados nas imagens selecionadas para teste ver Figura 32.

Foram utilizadas medidas de orientação da matriz de co-ocorrência, com a utilização dos ângulos 0° , 45° , 90° , 135° e tomados os seus valores médios.

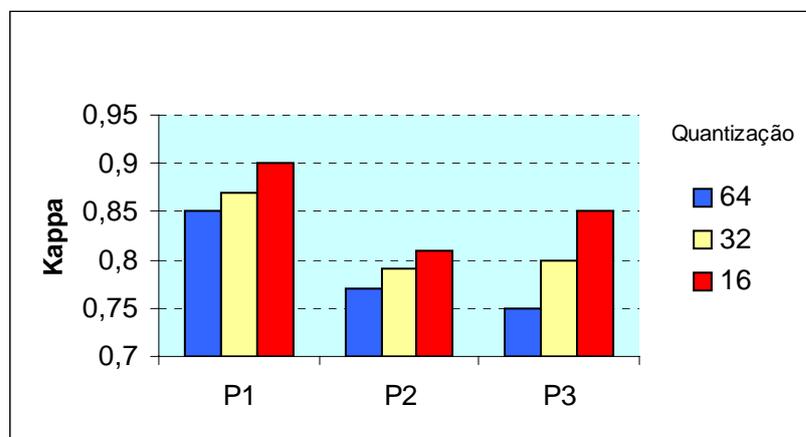


Figura 32 – Classificação das imagens de teste com diferentes níveis de quantização, utilizando GLCM.

5.3.1 Reduzindo o Espaço de Entrada de GLCM

Com o uso da técnica de redução da dimensão dos dados, a saber: PCA, ver seção 2.4.1, pode-se melhorar o desempenho dos dados extraídos de GLCM, como demonstrado na seção 3.2.1, aplicando-se PCA aos ângulos espaciais.

Na Tabela 4 e na Figura 33 podem ser visualizados um aumento de desempenho na classificação com o uso dessa técnica nas imagens de teste. Nessa análise foram utilizados níveis de quantização $N_g = 16$, distancia $d = 1$, variação de 95% no cálculo das componentes principais e as funções de características $F1$, $F2$, $F3$, $F4$, $F5$ e $F8$ ilustradas na Tabela 1.

Tabela 4 – Desempenho da utilização de PCA nos ângulos de GLCM em comparação com o uso de todos os ângulos e a média deles, nas imagens de teste.

Método	P1		P2		P3	
	Nº parâmetros	Kappa	Nº parâmetros	Kappa	Nº parâmetros	Kappa
Todos os ângulos	24	0,90	24	0,81	24	0,85
Média dos ângulos	6	0,87	6	0,79	6	0,83
PCA	15	0,90	18	0,81	13	0,85

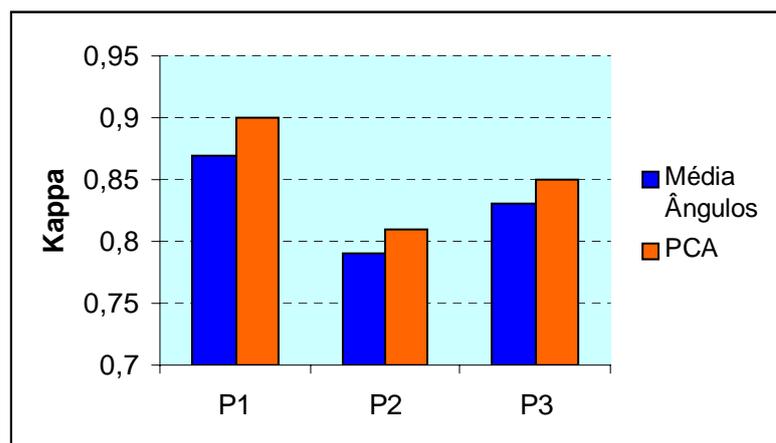


Figura 33 - Comparação da utilização de PCA nos ângulos de GLCM e a média dos ângulos nas imagens de teste.

Pode ser observado que aplicando-se PCA, a qualidade da classificação é igual a da classificação com todos os parâmetros, e possui melhor desempenho que com a utilização da soma dos ângulos da GLCM, além de reduzir o custo computacional envolvido.

Nas experiências realizadas anteriormente, foram utilizadas seis funções de GLCM, a saber, $\{F1, F2, F3, F4, F5 \text{ e } F8\}$ e seus respectivos ângulos devidamente analisados com o uso de PCA, porém existem muitas outras funções de GLCM que podem ser utilizadas na classificação, ver Tabela 1, porém deseja-se extrair somente um número mínimo de características de texturas usando-se essas funções, conforme mencionado no Capítulo 3. Na escolha desse um conjunto de características ótimo, torna-se necessário um modo de mensurar a capacidade desse conjunto em discriminar com bastante eficiência as classes de objetos envolvidas. Na seleção das melhores características de textura, obtidas das matrizes de co-ocorrência, para o processo de escolha, foi construído um classificador, baseado em rede neural probabilística, para um conjunto reduzido de características e escolhidas aquelas que produzem um melhor resultado em um conjunto separado de teste/validação, conforme visto

na seção 2.6. Nesta aproximação, o conjunto de características é escolhido de forma a adaptar-se ao classificador.

Uma vez definida a função de discriminação (a rede neural PNN), a ser utilizada para a seleção das melhores características, dentre todas envolvidas, é aplicado o algoritmo *plus l – take away r* (seção 2.6.2) para essa escolha, sendo computacionalmente satisfatório para a maioria das aplicações. Por essa razão esse algoritmo foi escolhido para a seleção de características. A função de discriminação foi construída a partir de um classificador que utiliza uma rede neural PNN, do mesmo tipo utilizada no módulo classificador principal. O processo descrito pode ser visualizado em etapas na Figura 32.

Na Tabela 5 são mostrados os resultados obtidos após o uso do classificador proposto, com o algoritmo de escolha, mostrando a quantidade de funções de GLCM selecionadas de forma automática para o grupo de texturas em análise. Pode ser visto que a escolha das funções, pelo algoritmo de seleção, que possuem maior poder discriminatório das texturas em análise, possibilitou uma melhora de até 3,5% em comparação ao uso de muitas funções.

Na Tabela 6 são mostradas as funções selecionadas, pelo algoritmo, para a imagem de teste P3, sendo essas funções as que melhores possibilitam (dentro da capacidade do algoritmo) a correta classificação temática da imagem.

A Tabela 7 mostra os resultados, em termos de coeficiente de Kappa, das imagens de teste mostradas na Figura 26 e Figura 27, incluindo a utilização das imagens coloridas.

O resultado final da classificação pode ser visto na Figura 35 e na Figura 36 que mostram as imagens resultantes com a aplicação da metodologia proposta.

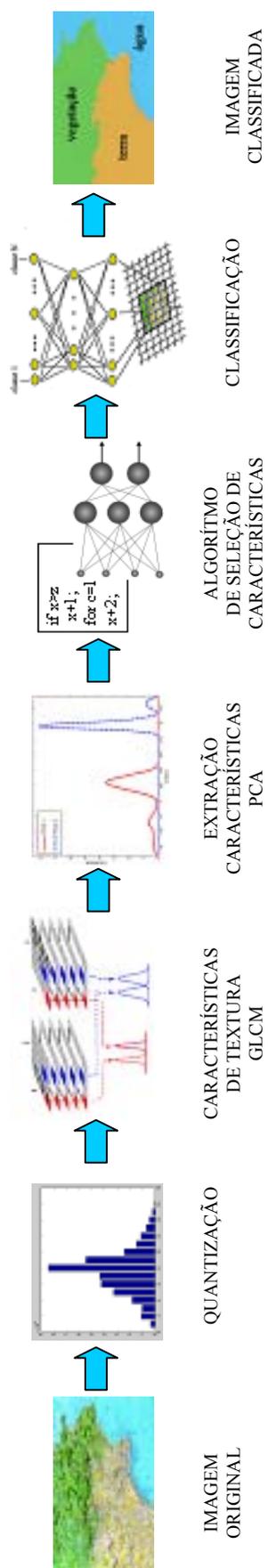


Figura 34 – Etapas do processo de classificação de imagem com o uso de GLCM descrito.

Tabela 5 – Resultados obtidos com a redução de funções de GLCM

Imagem	Nº parâmetros (após PCA)	Kappa	Nº parâmetros selecionados	Kappa
P1	25	0,92	4	0,94
P2	36	0,83	21	0,86
P3	29	0,88	15	0,89

Tabela 6 – Parâmetros selecionados pelo algoritmo para a imagem de teste P3

Função	Nº componentes PCA
<i>F3</i>	2
<i>F5</i>	3
<i>F6</i>	1
<i>F7</i>	1
<i>F8</i>	1
<i>F9</i>	1
<i>F13</i>	2
<i>F14</i>	2
<i>F21</i>	2

Tabela 7 – Resultado comparativo da classificação, segundo o método proposto, nas imagens de teste.

Imagem	Kappa
P1	0,94
P2	0,86
P3	0,89
S2	0,82

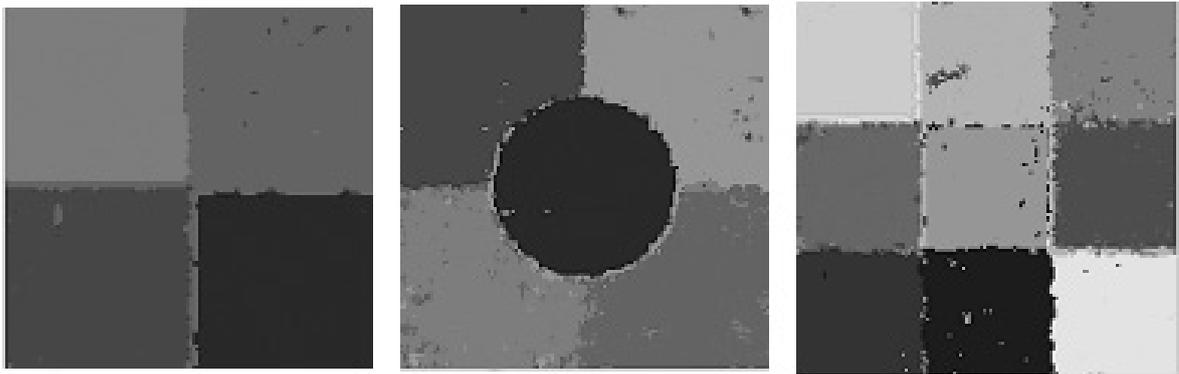


Figura 35 – Resultado da classificação das imagens de teste P1, P2, P3, com a utilização do método GLCM.

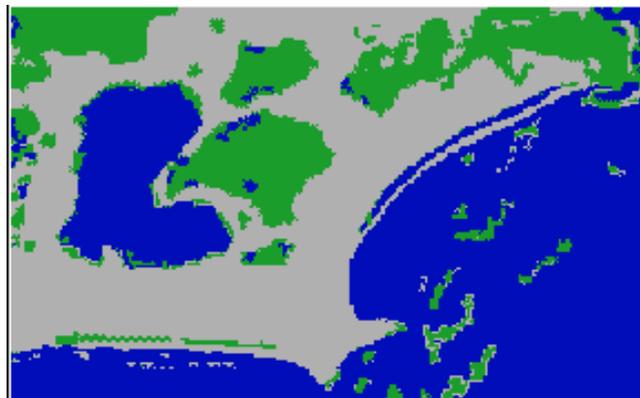


Figura 36 – Resultado da classificação, com três classes temáticas, da imagem de teste S2, com o uso de GLCM.

5.4 Avaliação e Testes com SGF

O processo de extração com SGF começa pela escolha do limiar de separação das imagens, de forma a produzir um número de planos binários, de acordo com a equação (10), vista na seção 3.3. O número de regiões de valores 1 e 0 em cada plano informa as medidas geométricas.

O método utilizado neste trabalho usa somente 63 planos, ou imagens binárias, com limiares $\tau = 4, 8, 12, \dots, 252$, de forma a reduzir custos computacionais, semelhantes aproximações foram feitas por Chen et al. (1995) e Runnacles & Nixon (1996). A Figura 37 mostra algumas imagens de planos de bits, da imagem de teste P3, com valores de limiar $\tau = 4, 12, 24, 32, 44, 64, 96, 120, 196, 232$.

Após a aplicação das fórmulas citadas na seção 3.3, são calculadas as regiões 4-conectadas das imagens de teste. Utilizando-se a mesma janela deslizante descrita em 5.2.1, com tamanho de 10×10 pixels, teremos um total de 16 características amostradas.

O classificador utilizado é a rede PNN de três camadas, contendo um total de 16 neurônios na camada de entrada (20 quando da utilização de cores na classificação) e a mesma quantidade de neurônios de saída quantas forem às classes da imagem de teste, ou seja, 3, 4, 5 ou 9 neurônios. Os parâmetros extraídos das imagens de teste foram normalizados antes de serem aplicados na rede neural. O processo descrito é mostrado na

Figura 38 e o resultado da classificação das imagens de teste mostradas na Figura 26 e Figura 27 pode ser visto na Tabela 8 e as imagens resultantes da classificação na Figura 40 e Figura 41.

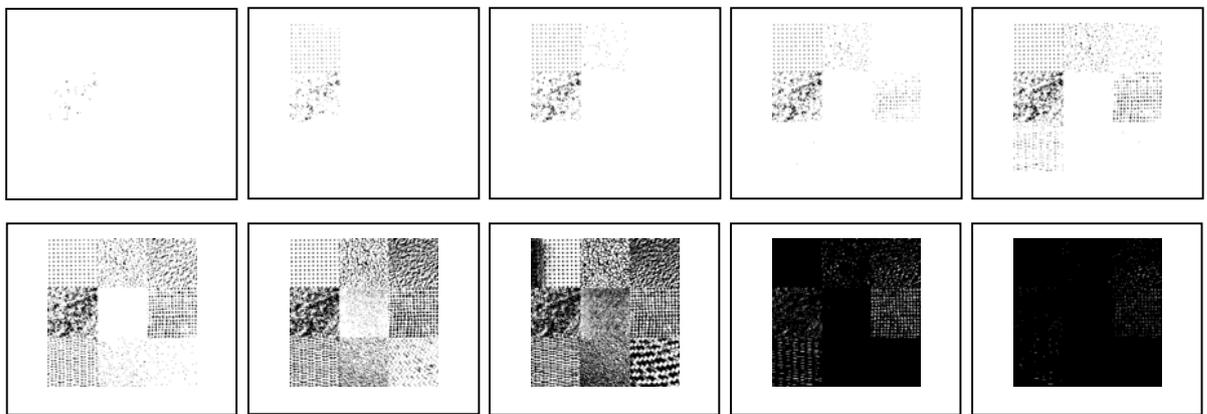


Figura 37 – Planos de bits da imagem de teste P3 para valores de $\tau = 4, 12, 24, 32, 44, 64, 96, 120, 196, 232$.

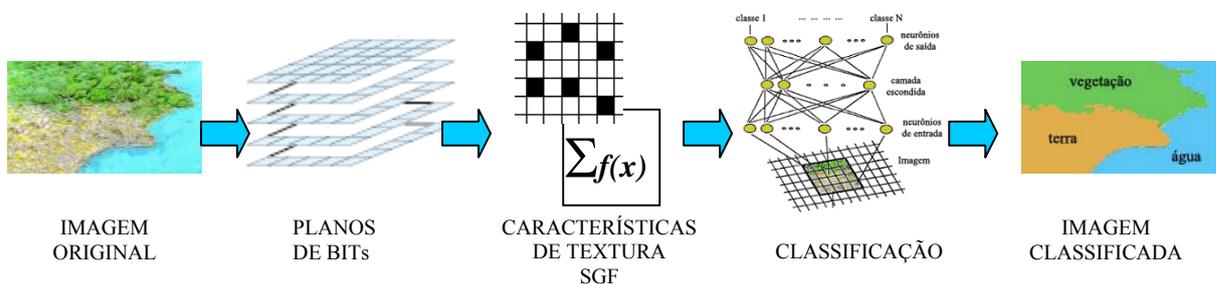


Figura 38 – Etapas do processo de classificação com o uso de SGF.

Tabela 8 – Resultados da aplicação de SGF

Imagem	Kappa
P1	0,95
P2	0,88
P3	0,91
S2	0,85

5.4.1 Reduzindo o Espaço de Entrada de SGF

A seguir foram realizados experimentos com o uso de PCA e o algoritmo de seleção de características, exposto na seção anterior (ver também seção 2.6.2), de forma a avaliar o uso dos mesmos nos parâmetros extraídos das imagens de teste.

Ambos os testes foram realizados individualmente, de forma a verificar a eficiência de cada um.

Na Tabela 9 podem ser visualizados os resultados com a aplicação de PCA, com variação de 95% no cálculo das componentes principais, nos 16 parâmetros de SGF.

Tabela 9 – Utilização de PCA nos parâmetros de SGF

Imagem	Nº parâmetros (após PCA)	Kappa
P1	10	0,95
P2	11	0,89
P3	14	0,91
S2	10	0,86

De forma semelhante descrita na seção 5.3.1 quando do uso de GLCM, foi utilizado o algoritmo de seleção *plus l – take away r* juntamente com a função de discriminação (uma rede neural PNN) de forma a selecionar os melhores parâmetros, dentre os 16 envolvidos.

Os resultados são mostrados na Tabela 10 e na Figura 39 pode ser visto uma comparação dos métodos de melhoria de desempenho, com a redução de parâmetros, utilizando SGF.

De acordo com os experimentos o melhor resultado foi obtido com o uso do algoritmo de seleção dos melhores parâmetros.

Os melhores resultados de classificação podem ser vistos na Figura 40 e na Figura 41, que mostram os resultados na classificação das imagens de teste, com o algoritmo de seleção utilizado.

Tabela 10 – Resultados obtidos com a redução do espaço de entrada com SGF

Imagem	Nº parâmetros selecionados	Kappa
P1	8	0,95
P2	10	0,90
P3	12	0,91
S2	9	0,87

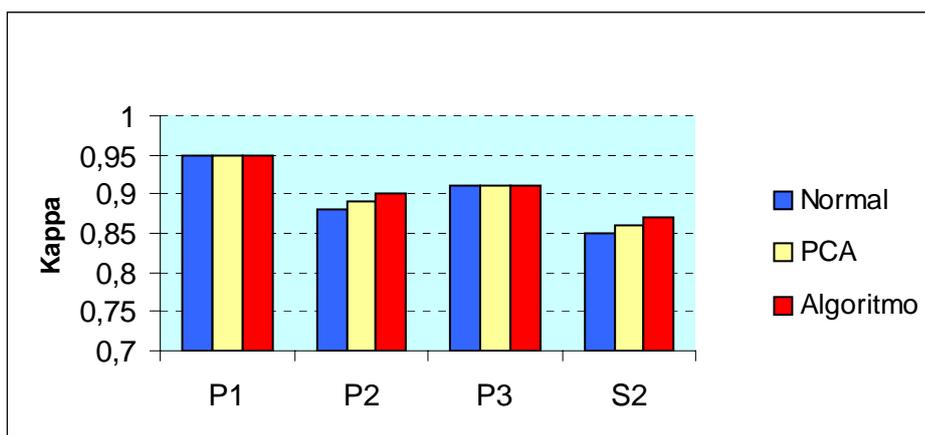


Figura 39 – Comparação dos métodos de redução de dimensão dos parâmetros de SGF.

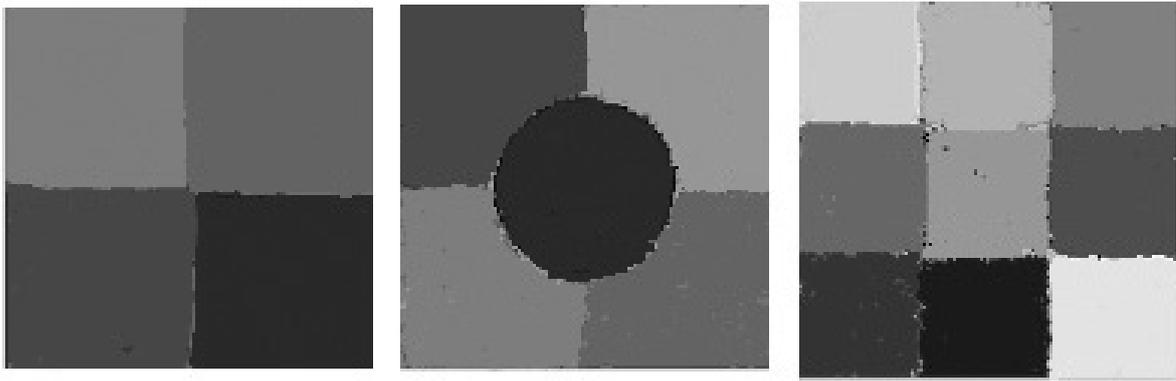


Figura 40 – Resultado da classificação das imagens de teste P1, P2, P3, com a utilização do método SGF e o algoritmo de seleção.



Figura 41 – Resultado da classificação, com três classes temáticas, da imagem de teste S2 com o uso de SGF e o algoritmo de seleção.

5.5 Comparação dos Métodos GLCM e SGF

Nos experimentos de classificação utilizados, SGF possui um desempenho superior a GLCM com um erro de classificação de até 6,1%. Os resultados das comparações entre os métodos podem ser vistos na Tabela 11, e na Figura 42 que mostram os resultados comparativos.

Tabela 11 – Resultados comparativos (índice de Kappa) entre GLCM e SGF.

Imagem	Método	
	GLCM	SGF
P1	0,94	0,95
P2	0,86	0,90
P3	0,89	0,91
S2	0,82	0,87

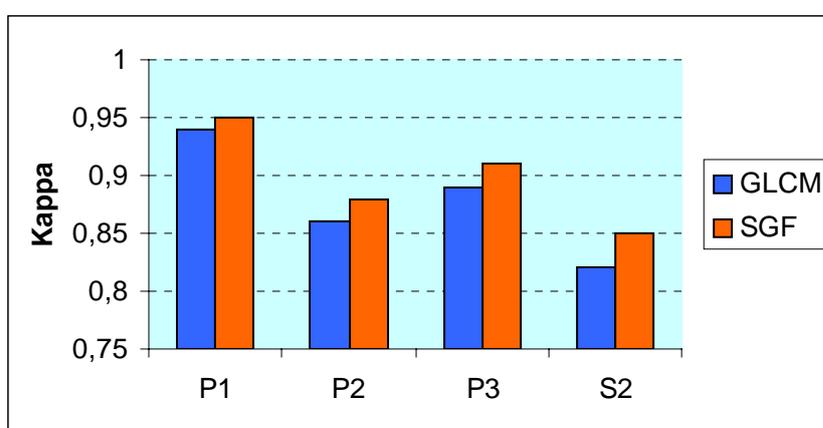


Figura 42 – Comparação entre os métodos GLCM e SGF nas imagens de teste.

5.6 Uso de Cores

Neste trabalho optou-se pelo método de extração de informações de cores em separado das informações de textura, conforme visto na seção 2.3.2.1. As informações de cores foram extraídas de acordo com o método disposto na seção 2.3.3. As imagens foram mapeadas para o espaço de cores HSV e extraídas informações de média e desvio padrão das informações de cromaticidade, componentes H (matiz) e S (saturação), das imagens e posteriormente aplicadas ao classificador (rede neural probabilística), juntamente com as informações de textura, componente V (valor – intensidade), provenientes do método utilizado, seja GLCM ou SGF. O processo descrito pode ser visto na Figura 43.

Os resultados da utilização de cores podem ser visualizados na Tabela 12 e na Figura 44, que mostram os resultados dos testes com GLCM e SGF. Resultados comparativos mostram que utilizando informações de cores consegue-se um melhor desempenho de classificação de até 3,7% com o uso de GLCM e de até 2,3% com SGF.

As imagens mostradas na Figura 45 e Figura 46, representam a classificação, das imagens de teste (P1c, P2c, P3c), com a utilização dos métodos GLCM e SGF aplicadas ao classificador PNN. Os resultados com a imagem de satélite S1, são mostrados na Figura 47 e Figura 48, respectivamente para GLCM e SGF. A Tabela 11 mostra uma comparação, em termos do índice de Kappa, entre os métodos utilizados.

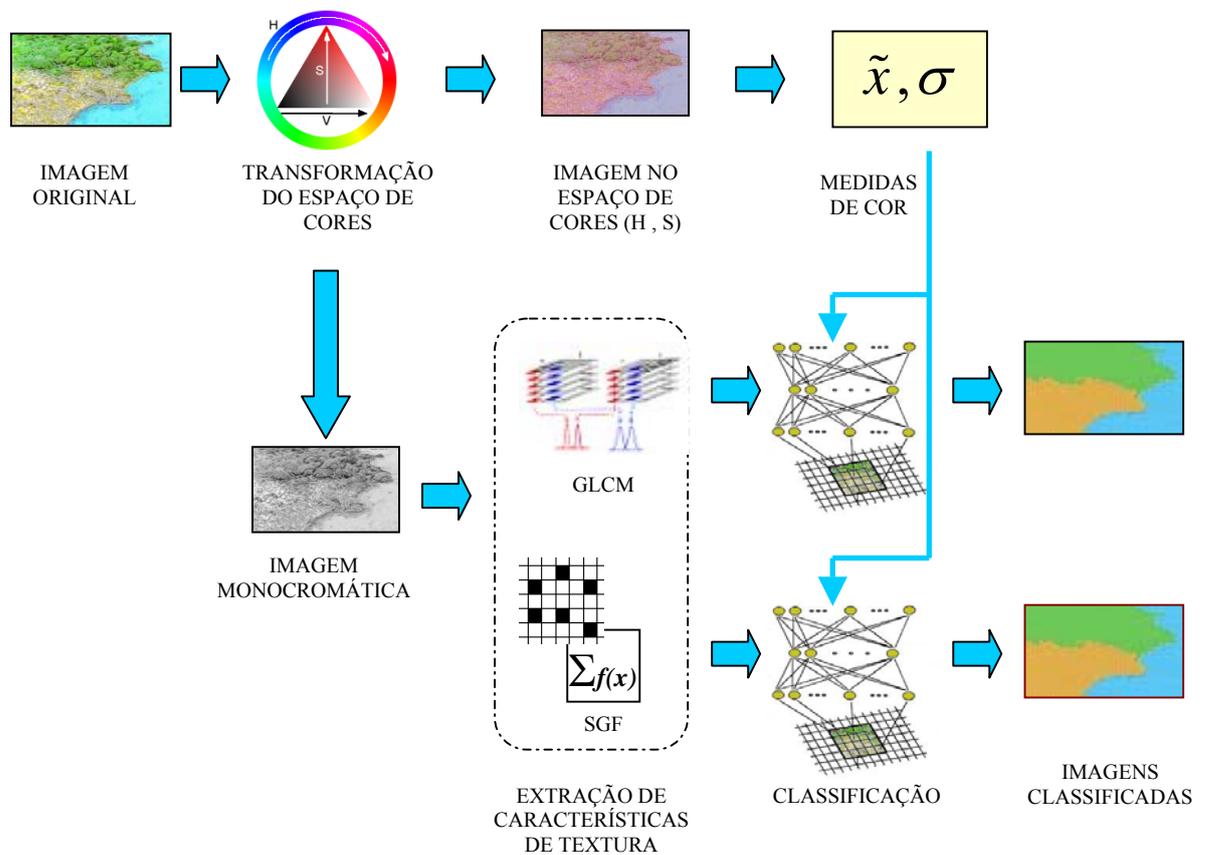


Figura 43 – Etapas do processo de classificação com o uso de cores.

Tabela 12 – Resultados com o uso de informações de cores em GLCM e SGF.

Imagem	Kappa	
	GLCM	SGF
P1c	0,95	0,96
P2c	0,87	0,91
P3c	0,91	0,92
S1	0,85	0,89

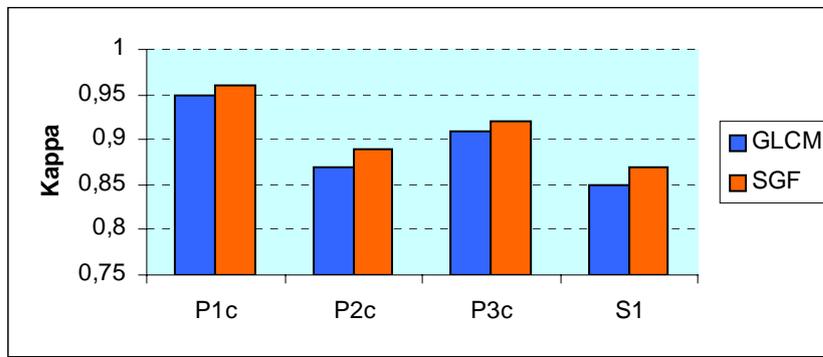


Figura 44 - Comparação entre os métodos GLCM e SGF com o uso de imagens coloridas.

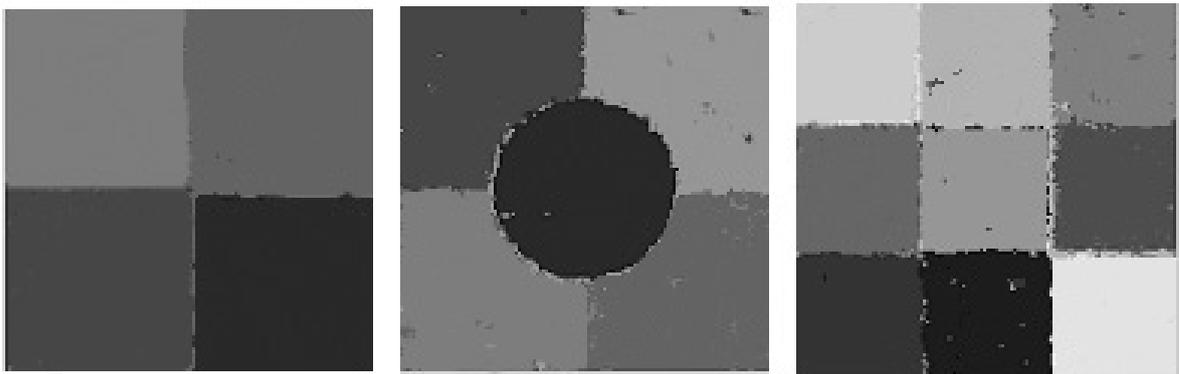


Figura 45 – Resultado da classificação das imagens de teste P1c, P2c, P3c, com a utilização do método GLCM.

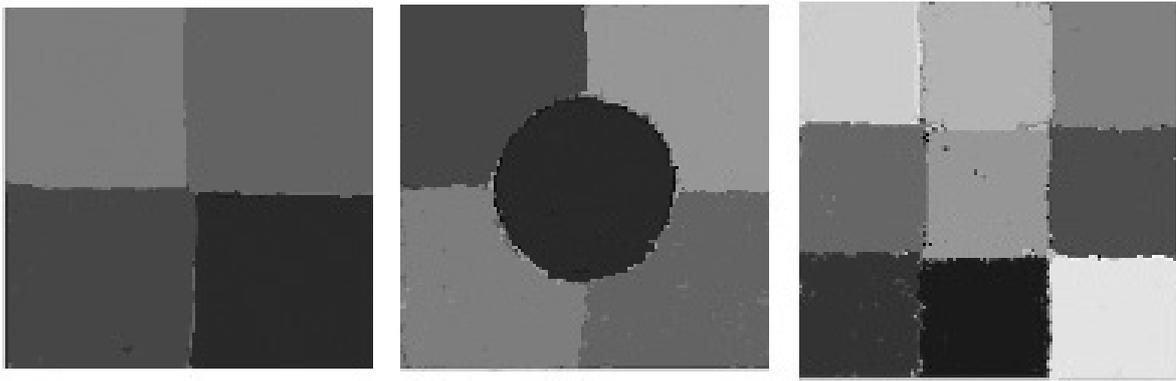


Figura 46 – Resultado da classificação das imagens de teste P1c, P2c, P3c, com a utilização do método SGF.



Figura 47 – Resultado da classificação, com três classes temáticas, da imagem de teste S1 com o uso de GLCM.



Figura 48 – Resultado da classificação, com três classes temáticas, da imagem de teste S1, com o uso de SGF.

Capítulo 6

Conclusões

Este trabalho propõe a avaliação de duas metodologias para classificação supervisionada de imagens digitais baseadas em características de textura. Nessa abordagem, a textura é descrita em termos de medidas estatísticas calculadas através de matrizes de co-ocorrência e descritores estatísticos de formas geométricas. Além disso, outras técnicas foram avaliadas, como a utilização de cores, no processo de classificação temática das imagens envolvidas e o uso de classificadores neurais, como a rede perceptron multicamadas e a rede probabilística.

GLCM tem demonstrado ser uma técnica de caracterização de texturas ainda muito eficiente, considerando texturas como uma distribuição estatística de relações espaciais de níveis de cinza em imagens. A quantização uniforme foi utilizada como um pré-processamento necessário para o cálculo das matrizes de co-ocorrência. Métodos de seleção de características foram usados de forma a selecionar os melhores parâmetros obtidos com GLCM de forma a evitar parâmetros redundantes.

O algoritmo de GLCM é custoso em termos computacionais, durante a obtenção das matrizes e das diversas funções calculadas, o uso do discriminador proposto aliado ao algoritmo de seleção permite uma redução no número de funções de GLCM que possuem um alto poder discriminatório das texturas em análise. Como resultado o uso dessa técnica permite reduzir erros de classificação oriundos da utilização de diversas funções com baixo poder discriminatório. Foi demonstrada essa melhoria de desempenho na comparação do uso de GLCM isolada e com o algoritmo de seleção. A aplicação desse método não está limitada ao uso de GLCM, podendo ser estendida para qualquer método de análise onde uma série de funções de discriminação necessitem ser escolhidas.

O conjunto de funções de distribuição estatística de formas geométricas representa um novo método para caracterização de imagens baseadas em texturas, apresentando resultados superiores em comparação com GLCM nas imagens utilizadas. Entretanto, o uso de SGF embora superior ao GLCM, não apresentou desempenho muito significativo ao uso de GLCM aliada ao algoritmo de seleção proposto.

Nas duas metodologias, optou-se por empregar a rede neural probabilística, por apresentar desempenho superior a da MLP, principalmente devido a uma melhor estimativa de parâmetros das classes envolvidas e também ao menor tempo de treinamento, tempo esse bastante inferior ao tempo de treinamento da MLP. Resultados semelhantes foram obtidos por outros autores, citados nesse trabalho, utilizando esse tipo de rede neural para classificação de imagens.

O uso de imagens coloridas aliadas às características de texturas possibilita aumentar o desempenho de classificação de imagens, embora o uso de texturas em imagens monocromáticas tenha seu uso justificado quando da utilização para classificação temática de imagens oriundas de radar.

Os resultados experimentais demonstraram o desempenho dos métodos citados, permitindo sua aplicação em uma variedade de imagens, podendo ser utilizadas imagens multiespectrais de sensoriamento remoto.

Como trabalho futuro, devido aos cálculos computacionais custosos na extração das características de textura, a computação paralela aparece como uma alternativa viável para a obtenção do desempenho necessário para a solução desses problemas. Podendo-se através da utilização de clusters, ou com as novas tendências de utilização de processadores multi-núcleo, uma alternativa para a obtenção deste desempenho.

Referências

- ANDREWS, H. C., “Introduction to Mathematical Techniques in Pattern Recognition”. *Wiley-Interscience, John Wiley & Sons, 1972.*
- ANDREY, P.; TARROUX, P., “Unsupervised segmentation of Markov random field modeled textured images using selectionist relaxation”. *Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume: 20 Issue: 3 Mar 1998 Page(s): 252-262.*
- ARROWSMITH, M. J., VARLEY, M. R., PICTON, P.D., HEYS, J. D., “Hybrid Neural Network System for Texture Analysis”. *Image Processing and its Applications, Conference Publication No. 465 IEE 1999.*
- ARTUR, J. S., “Reconhecimento de Padrões Usando Indexação Recursiva”. *Universidade Federal de Santa Catarina, 1999.*
- BARBER, D. G. & LEDREW E. F., “Sar sea ice using texture statistics: a multivariate approach”. *Photogrammetric Engineering & Remote Sensing”. V. 57, n. 4, p. 385-395, abr. 1991.*
- BELLMAN, R., “Adaptive Control Processes: A Guided Tour”. *Princeton University Press, 1961.*
- BEZDEK, C. J. e PAL, S. K., “*Fuzzy Models for Pattern Recognition*”. *IEEE Press, New York, 1992.*
- BIN TIAN, SHAIKH, M.A., AZIMI-SADJADI, M.R., HAAR, T.H.V. and REINKE, D.L., “A study of cloud classification with neural networks using spectral and textural features” *Neural Networks, IEEE Transactions on Volume: 10 Issue: 1 Jan 1999 Page(s): 138-151.*
- BISHOP, C. M., “Neural Networks for Pattern Recognition”. *Oxford University Press Inc., USA, NY, New York, 1995.*
- BOVIK, A.C., CLARK, M., and GEISLER, W.S., “Multichannel texture analysis using localised spatial filters”, *IEEE Trans. Pattern Anal. Mach. Intell.*, 1990, 12, (1), pp. 55-73.
- BRIGGS, J.M., and NELLIS, M.D., “Seasonal Variation of Heterogeneity in the Tallgrass Prairie: A Quantitative Measure Using Remote Sensing”. *Photogrammetric Engineering and Remote Sensing. 57 (4): 407-411, 1991.*
- CAELLI, T. and REYE, D., “On the classification of image regions by colour, texture and shape”. *Pattern Recognition 26 (4) 461–470, 1998.*
- CHELLAPPA. R.. and CHATTERJEE. S., “Classification of textures using Gaussian Markov random fields”, *IEEE Trans.*, 1985, *ASSP-33, (4), pp. 959-963.*
- CHEN, Y. Q., NIXON, M. S. and THOMAS, D. W., “Statistical Geometric Features for Texture Classification”. *Pattern Recognition 28(4) 537—552, 1995.*
- CHRISTODOULOU, C.I., MICHAELIDES, S.C., PATTICHIS, C.S. and KYRIAKOU, K., “Classification of satellite cloud imagery based on multi-feature texture analysis and neural networks”. *Image Processing, 2001. Proceedings. 2001 International Conference on Volume: 1 2001 Page(s): 497-500 vol.1*
- CICHOCKI A., UNBENHABEN, R., “Neural Networks for Optimization and Signal Processing”. *John Wiley & Sons inc, 1993.*
- CONGALTON, R. G.; GREEN, K., “A practical look at sources of confusion in error matrix generation”. *Photogrammetric Engineering and Remote Sensing, v. 59, n. 5, p. 641-644, 1993.*

- CONNERS, R. W., TRIVEDI, M. M. and HARLOW, C. A., "Segmentation of a High-Resolution Urban Scene using Texture Operators". *Computer Vision, Graphics and Image Processing*, 25:273-310, 1984.
- CONNERS, R.W. & HARLOW, C. A., "Equal probability quantizing and texture analysis of radiographic images". *Computer Graphics and Image Processing* 8, 447-463. 1978.
- CORDEIRO, F. M., "Reconhecimento e Classificação de Padrões de Imagens de Núcleos Linfócitos do Sangue Periférico Humano com a Utilização de Redes Neurais Artificiais". *Universidade Federal de Santa Catarina*, 2002.
- CRÓSTA, A. P., "Processamento digital de imagens de sensoriamento remoto". *Campina. IG/UNICAMP*. 1992.
- DAVIS, L. S., JOHNS, S. A., AGGARWAL, J. K., "Texture analysis using generalized cooccurrence matrices". *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1(3)*, 251-259. 1979.
- DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B., "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B*, vol. 39, pp. 1-38, 1977.
- DENG, Y.; MANJUNATH, B.S., "Unsupervised segmentation of color-texture regions in images and video". *Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume: 23 Issue: 8 Aug 2001 Page(s): 800-810*.
- DONALD, F. Specht, "Probabilistic Neural Networks", *Neural Networks*, vol. 3, pp. 109-118, 1990.
- DUBUISSON, M. P and GUPTA, J. A. "Color and texture fusion: application to aerial image segmentation and GIS updating". *Image Vision Comput.* 1, 823-832, 2000.
- GALLOWAY, M. M., "Texture analysis using gray level run lengths". *Computer Graphics and Image Processing* 4, 172-179, 1995.
- GLOTFELTY, J. E., "Automatic Selection of Optimal Window Size and Shape for Texture Analysis". *Master of Arts in Geography, West Virginia University*, 1999.
- GONZALEZ, R. C., WOODS, R. E., "Digital Image Processing", *Prentice Hall, New Jersey*, 2002.
- GOOL, L. V., DEWAELE, P., OOSTERLINCK, A. , "Texture analysis anno 1983". *Computer Vision, Graphics, and Image Processing* 51, 70-86. 1990.
- HARALICK, R. M. & SHAPIRO L. G., "Texture. In: Computer and robot vision". v. 1. *New York, USA : Addison-Wesley Publishing Company*. p. 453-494, 1992.
- HARALICK, R. M., "Statistical and Structural Approaches to Texture", *Proceeding of the IEEE*, vol. 67, no. 5, May 1979, pp 786-804.
- HARALICK, R. M., Shunmugam, K. and Dinstein, I., "Textural Features for Image Classification", *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, no. 6, Nov 1973.
- HUANG, J. et.al., "Image indexing using color correlogram". *Proceeding of Computer Vision and Pattern Recognition*, pp.762-768, 1997.
- HUAWU, Deng and CLAUSI, D.A., "Advanced Gaussian MRF rotation-invariant texture features for classification of remote sensing imagery". *Computer Vision and Pattern Recognition*, 2003. *IEEE Computer Society Conference on Volume: 2 18-20 June 2003 Page(s): II- 685-90 vol.2*

- KASHYAP, R. L., CHELLAPPA, R. & KHOTANZAD, A., “Texture classification using features derived from random field models”. *Pattern Recognition Letters 1*, 43-50. 1982.
- KIM, K.H.; SHARIF, B.S.; CHESTER, E.G., “Unsupervised texture analysis using a robust stochastic image model”. *Image Processing and Its Applications, 1997., Sixth International Conference on Volume: 2 14-17 Jul 1997 Page(s): 613-617 vol.2.*
- KITTLER, J., “Feature set search algorithms in Pattern Recognition and Signal Processing”. *Sijthof and Noordhoff, The Netherlands, 1978.*
- MÄENPÄÄ, T., PIETIKÄINEN, M., “Classification with color and texture: jointly or separately?”. *Pattern Recognition (37), No. 8, August 2004, pp. 1629-1640.*
- MARCEAU, D.J., Howarth, P.J., Dubois, J.M. and Gratton, D.J., “Evaluation of the Grey-Level Co-occurrence Matrix Method for Land-Cover Classification Using SPOT Imagery”. *IEEE Transactions on Geoscience and Remote Sensing. 28 (4): 513-519, 1990.*
- MARTIN, Erik, “Pap-Smear Classification”, *Technical University of Denmark, september, 2003.*
- MUSCI, Marcelo. Z. A., “Um Método para Classificação Supervisionada de Imagens de Sensoriamento Remoto Utilizando Rede Neural Probabilística e Características de Textura”, *Revista de Ciência & Tecnologia, Universidade Iguazu, vol. 6 – n° 1- Jun/2006 p. 69-80.*
- NATURAL RESOURCES CANADA. Disponível em: <http://www.ccrs.nrcan.gc.ca>.
- NISHIDA, W. and BASTOS, L. C., “Classificação de Imagens de Sensoriamento Remoto Utilizando uma Rede Neural Artificial com Função de Base Radial”, *Anais IX Simpósio Brasileiro de Sensoriamento Remoto, Santos, Brasil, 11-18 setembro 1998, INPE, p. 991-1001.*
- O'CALLAGHAN, R.J. and BULL, D.R., “Combined morphological-spectral unsupervised image segmentation”. *Image Processing, IEEE Transactions on Volume: 14 Issue: 1 Jan. 2005 Page(s): 49- 62.*
- OJALA, T., PIETIKAINEN, M. & HARWOOD, D., “Quantitative description of chromatin structure during neoplasia by the method of image processing”. *Pattern Recognition 29(1), 51-59. 1996.*
- PACHECO, Admilson e RIBAS, Nilton de Souza, “Sensoriamento Remoto Aplicado ao Uso da Terra”. *Revista da Comissão Brasileira de Geodésia. Geodésia online – 4/1998 (ISSN 1415-1111).*
- PANDYA, A., MACY, R.B., “Pattern Recognition with Neural Networks in C++”. *CRC Press, 1995.*
- PASCHOS, G. and PETROU, M., “Histogram ratio features for color texture classification”. *Pattern Recogn. Lett. 24 309–314, 2003.*
- PEDDLE, D.R. e FRANKLIN, S. E., “Image Texture processing and data integration for surface pattern discrimination”. *Photogrammetric Engineering & Remote Sensing. V. 57, n. 4, p. 413-420, abr. 1991.*
- PITOT, A., TOGA, A.W., AYACHE, N. and THOMPSON, P., “Texture based MRI segmentation with a two-stage hybrid neural classifier”. *Neural Networks, 2002. IJCNN '02. Proceedings of the 2002 International Joint Conference on Volume: 3 2002 Page(s): 2053-2058.*
- POSTON, W. L., MARCHETTE, D. J., “Recursive dimensionality reduction using Fisher's linear discriminant”. *Pattern Recognition, v. 31, Issue: 7, p. 881-888, Julho 1998.*

- RAGHU, P. P. and YEGNANARAYANA, B., “Supervised Texture Classification Using a Probabilistic Neural Network and Constraint Satisfaction Model”, *IEEE Transactions on Neural Network*, vol9, no. 3, May 1998.
- RAMAKRISHNAN, A.G., KUMAR, Raja S. and RAGHU, Ram H.V., “Neural network-based segmentation of textures using Gabor features”. *Neural Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on 2002* Page(s): 365- 374.
- ROSENFELD, A., YE-WANG, C. and WU, A., “Multispectral texture”. *IEEE Trans. Systems, Man, Cybern.* 12 (1) 79–84, 1982.
- RUNNACLES, B.S., NIXON, M.S., “Texture Extraction and Segmentation via Statistical Geometric Features”, *Image Processing, 1996. Proceedings., International Conference on Volume 3, Issue , 16-19 Sep 1996* Page(s):129 - 132 vol.3.
- SANTOS, A. P. Novo e LOMBARDO, M. A., “Metodologia de interpretação de dados de sensoriamento remoto e aplicações no uso da terra”. *Anais do Simpósio Brasileiro de Sensoriamento Remoto, São José dos Campos (SP), 1981, 7:172-175.*
- SCHOWENGERDT, R. A., “*Techniques for image processing and classification in remote sensing.*”. Orlando, Florida : Academic Pres. 1983.
- SCHWARTZ, W. R., “Método para classificação de imagens baseada em matrizes de co-ocorrência utilizando características de textura”. *III Colóquio Brasileiro de Ciências Geodésicas, Curitiba-PR, Brasil, pp. 1-11, 06 a 09 de Maio de 2003.*
- SING-TZE-BOW, “Pattern Recognition and Image Preprocessing”, *Marcel Dekker, 2 ed., New york, 2002.*
- SOUZA, J. A., “Reconhecimento de Padrões Usando Indexação Recursiva”. *Tese de doutorado, UFSC – 1999, Florianópolis.*
- SUN, C. & WEE, W. G., “Neighboring gray level dependence matrix for texture classification”. *Computer Vision, Graphics and Image Processing*, 23, 341-352. 1983.
- TAN, T. , KITTLER, J., “Colour texture classification using features from colour histogram”. *Proceedings of the 8th Scandinavian Conference on Image Analysis, Vol. 2, Tromso, Norge, 1993, pp. 807–813.*
- TATIBANA, C. Y. & KAETSU D. Y., Homepage de Redes Neurais, <http://www.din.uem.br/ia/neurais/>
- THEODORIDIS, Sergios and KOUTROUMBAS, Konstantinos, “Pattern Recognition”, *Elsevier, 2 ed., 2003.*
- UNSER, M., “Sum and difference histograms for texture classification”. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8(1), 118-125. 1986.*
- UNSER, M., “Texture classification and segmentation using wavelet frames”, *IEEE Trans. Image Process.*, 1995, 4, (11), pp. 1545 -1560.
- VISA, A., “Texture classification and segmentation based on neural network methods”. *Espoo, Finland. Tese (Doutorado em Tecnologia) Helsinki University of Technology. 1990.*
- WALKER, F. R., “Adaptive Multi-Scale Texture Analysis With Application to Automated Cytology”. *Doctor Thesis, The University of Queensland. 1997.*
- WEBB, Adrew R., “Statistical pattern Recognition”. *John Wiley & Sons, 2 ed. Malvern, UK,2002.*

WESZKA, J. S., DYER, C. R. and ROSENFELD, A., "A comparative study of texture measures for terrain classification". *IEEE Transactions on Systems, Man, and Cybernetics SMC-6(4)*, 269-285. 1976.

WU, C. M. & CHEN, Y. C., "Statistical feature matrix for texture analysis". *Computer Vision, Graphics and Image Processing*, 54, 407-419. 1992.

YIU, K.K., MAK, M.W. and LI, C.K., "Gaussian mixture models and probabilistic decision-based neural networks for pattern classification: A comparative study," *Neural Computing and Applications*, 8, 235-245, 1999.

YOGESAN. K., ALBREGTSEN, F. and DANIELSEN, H. E., "Gray level variance matrix: A new approach to higher order statistical texture analysis". *in Proceedings of the 3rd International Conference on Automation, Robotics and Computer Vision, Vol 2*, pp658-663. 1994.

Apêndice A

Fundamentação Estatística

A.1 Análise de Componentes Principais (PCA - *Principal Components Analysis*)

As técnicas de análise multivariada tem como característica comum à busca de “parâmetros-resumo”, que sintetizem a relação entre em determinado conjunto de variáveis. Dentro deste espírito, a técnica dos componentes principais busca imprimir um tratamento estatístico a um numero relativamente alto de variáveis heterogêneas, que possuam, porém, um grau considerável de aspectos comuns, isto é, com um elevado grau de correlação entre si. Desta forma o que se busca é condensar o conjunto inicial de muitas variáveis em um numero bem menor e conseguir uma pequena perda de informações.

Em termos práticos o que temos é o processo que se segue. Através de p variáveis originais, $X_1, X_2, X_3, \dots, X_p$ referentes a n parâmetros criam-se as variáveis $Z_1, Z_2, Z_3, \dots, Z_p$, os componentes principais, por meio da seguinte combinação linear:

$$\begin{aligned} Z_1 &= a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + \dots + a_{1p}X_p \\ Z_2 &= a_{21}X_1 + a_{22}X_2 + a_{23}X_3 + \dots + a_{2p}X_p \\ Z_3 &= a_{31}X_1 + a_{32}X_2 + a_{33}X_3 + \dots + a_{3p}X_p \\ Z_p &= a_{p1}X_1 + a_{p2}X_2 + a_{p3}X_3 + \dots + a_{pp}X_p \end{aligned} \tag{A.1}$$

Sujeito à restrição:

$$a_{i1}^2 + a_{i2}^2 + \dots + a_{ip}^2 = 1 \tag{A.2}$$

Tem-se matematicamente que esta combinação linear procede a transformação ortogonal do conjunto de variáveis originais correlacionadas em um novo conjunto de novas variáveis não correlacionadas. Importante destacar que as variáveis Z assim obtidas são capazes de, em ordem decrescente, resumir a variação dos dados originais, de modo que alguns poucos componentes são responsáveis pela maior parte da explicação total simplificando assim a análise:

$$Var(Z_1) \geq Var(Z_2) \geq \dots \geq Var(Z_p) \tag{A.3}$$

Teoricamente o número de componentes é sempre igual ao número de variáveis. O fato, porem, é que se temos um conjunto bem selecionado de variáveis de análise, a maioria das novas variáveis Z_i – os *componentes principais* – apresentam capacidade de explicação negligível, podendo ser então descartados, sem prejuízo da análise. As variâncias de Z_i são obtidas a partir dos autovalores da matriz de covariância ou de correlação das variáveis originais. Os autovetores, associados a cada autovalor, ordenados, fornecem os coeficientes –

a_i – para a equação acima, dos componentes principais, sendo o primeiro componente associado ao maior autovalor.

Os passos necessários para o cálculo das componentes principais pode ser descrito da seguinte forma:

1. Codificam-se as variáveis em X_1, X_2, \dots, X_p , e procede-se a sua standardização para que tenham média zero e variância constante igual a 1;
2. Calcula-se a matriz C de covariâncias, que corresponde a matriz de correlação, se as variáveis estão standardizadas;
3. Encontra-se os autovalores – $\lambda_1, \lambda_2, \dots, \lambda_p$ – e os correspondentes autovetores – a_1, a_2, \dots, a_p – sendo que estes últimos são os coeficientes do i -ésimo componente principal, e os primeiros (*eigenvalues*) são a variâncias dos mesmos;
4. Assume-se que apenas alguns poucos componentes Z_i , dado o elevado montante da variância dos dados que estes explicam, devem ser tomados como suficientes para orientar a análise satisfatoriamente.

A redução das variáveis a sua forma standardizada consiste simplesmente em retirar a média de cada observação de cada variável e em seguida dividir essa diferença pelo correspondente desvio padrão.

A definição a respeito do número de componentes a serem escolhidos como objeto de análise, cabe dizer que, cada trabalho que utilize o instrumental da análise de componentes principais deverá considerar com cuidado seus objetivos e as variáveis selecionadas para realizar tal corte metodológico. De modo geral, pode-se considerar que é desejável que o menor número de componentes explique a maior porcentagem da variância dos dados.

Considerando que o sistema de equações expresso em (A.1), pode ser reescrito na forma matricial como:

$$AX = \lambda X \quad (A.4)$$

ou

$$(A - \lambda I) = 0 \quad (A.5)$$

Onde:

A: matriz de coeficientes a_{ij}

X: matriz das variáveis

I: matriz identidade

0: vetor coluna de zeros

λ : autovalor ou raiz característica (escalar).

A matriz de correlação, considerando as variáveis normalizadas, é simétrica e assume a seguinte forma:

$$C = \begin{bmatrix} 1 & c_{12} & \cdots & c_{1p} \\ c_{21} & 1 & \cdots & c_{2p} \\ \vdots & \vdots & & \vdots \\ c_{p1} & c_{p2} & \cdots & 1 \end{bmatrix} \quad (\text{A.6})$$

Podemos obter os autovalores e autovetores da matriz C . O i -ésimo autovalor corresponde a variância do i -ésimo componente principal: $\text{var}(Z_i) = \lambda_i$. Os autovetores, obtidos através dos autovalores, correspondem aos pesos a serem dados às variáveis explicativas no processo de transformação que gera os índices Z_p (componentes principais).

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + a_{i3}X_3 + \dots + a_{ip}X_p \quad (\text{A.7})$$

Cumpramos destacar uma propriedade dos autovalores, qual seja, que a sua soma é igual ao traço da matriz de covariância. Ou seja:

$$\lambda_1 + \lambda_2 + \dots + \lambda_p = c_{11} + c_{12} + \dots + c_{pp} \quad (\text{A.8})$$

Dessa maneira, se c_{ii} é a variância de X_i e λ_i é a variância de Z_i , implica-se que o somatório das variâncias dos componentes principais é igual ao somatório das variâncias das variáveis originais.

A.2 Distância entre Distribuições

Essas medidas são usadas para se determinar o poder de discriminação das características de um conjunto de dados. Muitos métodos estatísticos foram propostos para o uso em reconhecimento de padrões, sendo os mais conhecidos os baseados no conhecimento das funções de densidade de probabilidade conjunta entre classes.

As medidas de distância de probabilidade utilizam a informação total sobre a estrutura das classes obtidas pelas densidades condicionais. A medida de distância J , satisfaz a seguinte condição:

1. $J = 0$ se as funções de densidade de probabilidade são iguais, $p(x | \omega_2)$;
2. $J \geq 0$;

3. J obtém seu máximo quando as classes são disjuntas, isto é, quando $p(x|\omega_1) = 0$ e $p(x|\omega_2) \neq 0$.

Muitas medidas satisfazendo essas condições foram propostas. Considerando duas distribuições sobrepostas com densidades condicionais $p(x|\omega_1)$ e $p(x|\omega_2)$. O erro de classificação, e , é dado por

$$e = \frac{1}{2} \left\{ 1 - \int |p(\omega_1|x) - p(\omega_2|x)| p(x) dx \right\} \quad (\text{A.9})$$

A integral na equação acima,

$$J_K = \int |p(\omega_1|x) - p(\omega_2|x)| p(x) dx \quad (\text{A.10})$$

é chamada de *distancia variacional de Kolmogorov* e possui importantes propriedades que estão diretamente relacionadas com o erro de classificação, mas pode ser usada para prover limites sobre o erro. Três dessas medidas são mostradas na Tabela A.1.

Tabela A.1 – Medidas de distância de probabilidade

Medidas	Forma matemática
Separação da média	$\frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} d(x_i, y_j), x_i \in \omega_A; y_j \in \omega_B$ <p>d qualquer distancia métrica</p>
Chernoff	$J_c = -\log \int p^s(x \omega_1) p^{1-s}(x \omega_2) dx$
Bhattacharyya	$J_B = -\log \int (p(x \omega_1) p(x \omega_2))^{\frac{1}{2}} dx$
Divergência	$J_D = \int [p(x \omega_1) - p(x \omega_2)] \log \left(\frac{p(x \omega_1)}{p(x \omega_2)} \right) dx$
Patrick-Fisher	$J_P = \left\{ \int [p(x \omega_1) p_1 - p(x \omega_2) p_2]^2 dx \right\}^{\frac{1}{2}}$

Uma das grandes desvantagens dos critérios envolvendo dependências de probabilidade é que se torna necessário estimar a função de densidade de probabilidade na integração numérica. Isso restringe seu uso em muitas situações práticas. Entretanto, sob certas suposições sobre as formas das distribuições, as expressões podem ser avaliadas analiticamente.

Primeiramente podemos considerar uma forma paramétrica específica para as distribuições, chamada de distribuição normal, com medias μ_1 e μ_2 e matrizes de covariância Σ_1 e Σ_2 . De acordo com essas suposições, as medidas de distância podem ser descritas da forma mostrada a seguir:

Chernoff

$$J_c = \frac{1}{2} s(1-s)(\mu_2 - \mu_1)^T [\Sigma_s]^{-1} (\mu_2 - \mu_1) + \frac{1}{2} \log \left(\frac{|\Sigma_s|}{|\Sigma_1|^{1-s} |\Sigma_2|^s} \right) \quad (\text{A.11})$$

onde $\Sigma_s = (1-s)\Sigma_1 + s\Sigma_2$ e $s \in [0,1]$. Para $s = 0,5$ temos a distância de Bhattacharyya.

Bhattacharyya

$$J_B = \frac{1}{4} (\mu_2 - \mu_1)^T [\Sigma_1 + \Sigma_2]^{-1} (\mu_2 - \mu_1) + \frac{1}{2} \log \left(\frac{|\Sigma_1 + \Sigma_2|}{2(|\Sigma_1| |\Sigma_2|)^{\frac{1}{2}}} \right) \quad (\text{A.12})$$

Divergência

$$J_D = \frac{1}{2} (\mu_2 - \mu_1)^T (\Sigma_1^{-1} + \Sigma_2^{-1}) (\mu_2 - \mu_1) + \text{Tr} \{ \Sigma_1^{-1} \Sigma_2 + \Sigma_2^{-1} \Sigma_1 - 2I \} \quad (\text{A.13})$$

Patrick-Fisher

$$J_P = (2\pi)^{-\frac{p}{2}} \left[|\Sigma_1|^{-\frac{1}{2}} + |2\Sigma_2|^{-\frac{1}{2}} - 2|\Sigma_1 + \Sigma_2|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mu_2 - \mu_1)^T (\Sigma_1 + \Sigma_2)^{-1} (\mu_2 - \mu_1) \right\} \right] \quad (\text{A.14})$$

Finalmente, se as matrizes de covariância são iguais, $\Sigma_1 = \Sigma_2 = \Sigma$, as distancias Bhattacharyya e divergência podem ser simplificadas para :

$$J_M = J_D = 8J_B = (\mu_2 - \mu_1)^T \Sigma^{-1} (\mu_2 - \mu_1), \quad (\text{A.15})$$

que recebe o nome de distância de Mahalanobis.

Naturalmente as médias e as matrizes de covariância não são conhecidas e precisam ser estimadas do conjunto de treino disponível.

As medidas de distancias de probabilidade podem ser estendidas para a forma multivariada pela avaliação de todas as distâncias de pares entre as classes,

$$J = \sum_{i=1}^C \sum_{j=1}^C p_i p_j I_{ij} \quad (\text{A.16})$$

onde J_{ij} é a medida escolhida das distancias medidas para as classes ω_i e ω_j .

A.3 Coeficiente de Kappa

A literatura existente sobre a estimação do erro da classificação é extensa e muitos procedimentos são propostos. Frequentemente, na prática, os resultados sobre as amostras usadas para treinamento e teste são expressos em uma forma tabular denominada matriz de classificação (ou matriz de confusão) que oferece a percentagem das amostras que foram correta e incorretamente classificadas.

A matriz de confusão de uma hipótese oferece uma medida efetiva do modelo de classificação, ao mostrar o número de classificações corretas *versus* as classificações preditas para cada classe, sobre um conjunto de exemplos T. A matriz de confusão pode ser vista na Figura A.2 e as classes são formadas de acordo com (A.17).

$$M(C_i, C_j) = \sum_{\{x, y\} \in T: y=C_i} \mathbb{1}_{\{h(x) = C_j\}} \quad (\text{A.17})$$

Classe	predita C_1	predita C_2	...	predita C_k
verdadeira C_1	$M(C_1, C_1)$	$M(C_1, C_2)$...	$M(C_1, C_k)$
verdadeira C_2	$M(C_2, C_1)$	$M(C_2, C_2)$...	$M(C_2, C_k)$
\vdots	\vdots	\vdots	\ddots	\vdots
verdadeira C_k	$M(C_k, C_1)$	$M(C_k, C_2)$...	$M(C_k, C_k)$

Figura A.2 – Matriz de confusão

O número de acertos, para cada classe, se localiza na diagonal principal $M(C_i, C_i)$ da matriz e os demais elementos $M(C_i, C_j)$, para $i \neq j$, representam erros na classificação. A matriz de confusão de um classificador ideal possui todos esses elementos iguais a zero uma vez que ele não comete erros.

A fim de se avaliar o desempenho da classificação, foi definido o coeficiente kappa, o qual é um parâmetro que pode ser extraído da matriz de classificação. O coeficiente é definido como:

$$\hat{K} = \frac{N \sum_{i=1}^r x_{ij} - \sum_{i=1}^r x_{i+} x_{+i}}{N^2 - \sum_{i=1}^r x_{i+} x_{+i}} \quad (\text{A.18})$$

onde r é o número de linhas ou de colunas da matriz de confusão; x_{ij} é o número de observações dos elementos da diagonal da matriz de confusão; $x_{i+} = \sum_j x_{ij}$ é a soma dos valores da linha i ; $x_{+i} = \sum_j x_{ji}$ é a soma dos valores da coluna i e N é o número total de observações.

Com frequência, na prática apresenta-se a seguinte forma equivalente para o coeficiente Kappa com propósitos computacionais:

$$\hat{K} = \frac{\theta_1 - \theta_2}{1 - \theta_2} \quad (\text{A.19})$$

onde

$$\theta_1 = \frac{1}{N} \sum_{i=1}^r x_{ii} \quad (\text{A.20})$$

$$\theta_2 = \frac{1}{N^2} \sum_{i=1}^r x_{i+} x_{+i} \quad (\text{A.21})$$

A variância do coeficiente pode ser estimada por:

$$\text{var}(\hat{k}) = \frac{1}{N} \left[\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2 - \theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4 - 4\theta_2^2)}{(1-\theta_2)^4} \right] \quad (\text{A.22})$$

onde θ_1 e θ_2 foram definidos acima e

$$\theta_3 = \frac{1}{N^2} \sum_{i=1}^r x_{ii}(x_{i+} + x_{+i}) \quad (\text{A.23})$$

e

$$\theta_4 = \frac{1}{N^3} \sum_{i=1}^r \sum_{j=1}^r x_{ij}(x_{j+} + x_{+i})^2 \quad (\text{A.24})$$

Quando todos os elementos da matriz de classificação fora da diagonal principal possuírem valores nulos, o coeficiente assumirá o valor 1, indicando um desempenho excelente. Também poderá assumir valores negativos indicando que o desempenho da classificação foi péssimo. A Tabela A.2 apresenta o desempenho da classificação em função do valor assumido pelo coeficiente Kappa.

Tabela A.2 – Desempenho da Classificação em Função do Coeficiente Kappa

Coeficiente Kappa	Desempenho da Classificação
$\hat{K} \leq 0$	Péssimo
$0 < \hat{K} \leq 0.2$	Mau
$0.2 < \hat{K} \leq 0.4$	Razoável
$0.4 < \hat{K} \leq 0.6$	Bom
$0.6 < \hat{K} \leq 0.8$	Muito Bom
$0.8 < \hat{K} \leq 1.0$	Excelente

Apêndice B

Classificadores Usados

B.1 Rede Neural Multicamadas (MLP – *Multilayer Perceptron*)

A micro-estrutura da rede MPL é definida pelas características de cada neurônio na rede, em particular pela sua função de ativação.

Na *Toolbox*¹⁹ de Redes Neurais do MATLAB® todos os neurônios de uma camada devem ter a mesma função de ativação. Em outros ambientes de simulação de redes neurais não há esta restrição, como p.ex. no NeuralWorks®.

Cada neurônio artificial, visto na Figura B.2 possui um número n de entradas cuja soma ponderada pelos pesos w_i passa pela função de ativação para gerar a saída do neurônio. Os pesos w_i são variáveis apenas na fase de treinamento. Depois desta fase o neurônio passa a ser uma função não-linear de $\mathfrak{R}^n \rightarrow \mathfrak{R}$. Uma entrada de polarização (*bias*) é freqüentemente utilizada, permitindo uma soma ponderada das entradas não nula quando a soma das entradas é zero. O parâmetro b , de deslocamento da função de ativação, torna-se assim desnecessário.

$$u = \sum_{i=1}^n w_i p_i - b, \quad a = f(u) = \begin{cases} +1 & \text{se } u \geq 0 \\ -1 & \text{se } u < 0 \end{cases} \quad (\text{B.1})$$

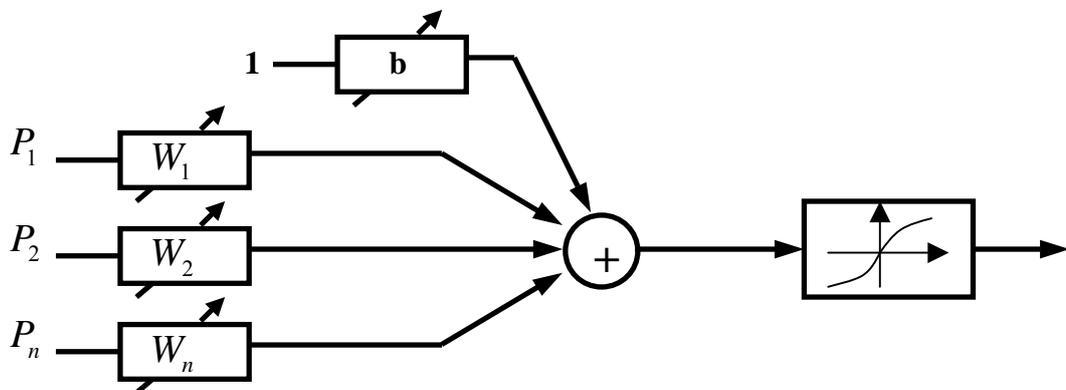


Figura B.2 – O neurônio artificial com entrada de polarização.

¹⁹ Conjunto de ferramentas de software para um propósito específico.

O Perceptron multicamadas pode ser visto na Figura B.2.

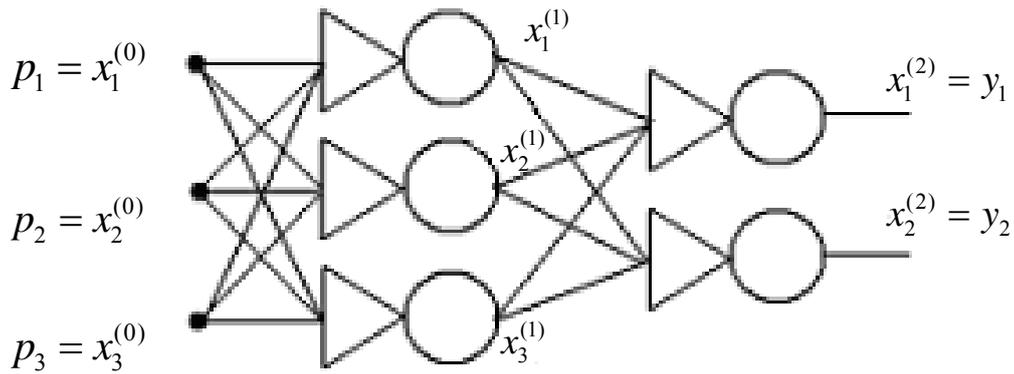


Figura B.2 – Perceptron Multicamadas (MLP)

Dinâmica: Elemento Processador (neurônio).

$$s_j^{(k)} = w_{0j}^{(k)} + \sum_i w_{ij}^{(k)} x_i^{(k-1)} \quad (\text{B.2})$$

$$x_j^{(k)} = f(s_j^{(k)}), \quad (\text{B.3})$$

com f contínua diferenciável.

Treinamento:

$$\varepsilon^2 = \sum_{j=1}^m (d_j - y_j)^2 \rightarrow \text{erro quadrático} \quad (\text{B.4})$$

$$w_j^{(k)} = (w_{0j}^{(k)}, w_{1j}^{(k)}, \dots, w_{mj}^{(k)}) \rightarrow \text{vetor de pesos do neurônio } j \quad (\text{B.5})$$

$$x_j^{(k-1)} = (1, x_{1j}^{(k-1)}, \dots, x_{nj}^{(k-1)}) \rightarrow \text{vetor de entradas do neurônio } j \quad (\text{B.6})$$

Gradiente instantâneo

$$\nabla_j^{(k)} = \frac{\partial \varepsilon^2}{\partial w_j^{(k)}} = \left[\frac{\partial \varepsilon^2}{\partial w_{0j}^{(k)}}, \frac{\partial \varepsilon^2}{\partial w_{1j}^{(k)}}, \dots, \frac{\partial \varepsilon^2}{\partial w_{mj}^{(k)}} \right] \quad (\text{B.7})$$

$$\nabla_j^{(k)} = \frac{\partial \varepsilon^2}{\partial w_j^{(k)}} = \frac{\partial \varepsilon^2}{\partial s_j^{(k)}} \frac{\partial s_j^{(k)}}{\partial w_j^{(k)}} \quad (\text{B.8})$$

$$\text{como } s_j^{(k)} = w_j^{(k)} x_j^{(k-1)} \quad (\text{B.9})$$

$$\frac{\partial s_j^{(k)}}{\partial w_{1j}^{(k)}} = x_j^{(k-1)} \quad (\text{B.10})$$

e portanto

$$\nabla_j^{(k)} = \frac{\partial \varepsilon^2}{\partial w_j^{(k)}} = \frac{\partial \varepsilon^2}{\partial s_j^{(k)}} x_j^{(k-1)} \quad (\text{B.11})$$

O erro quadrático é definido por

$$\delta_j^{(k)} = -\frac{1}{2} \frac{\partial \varepsilon^2}{\partial s_j^{(k)}} \quad (\text{B.12})$$

Para a camada de saída o erro derivativo quadrático é:

$$\delta_j^{(k)} = \frac{1}{2} \frac{\partial \sum_{i=1}^{N_k} (d_i - y_i)^2}{\partial s_j^{(k)}} = -\frac{1}{2} \frac{\partial \sum_{i=1}^{N_k} (d_i - f(s_j^{(k)}))^2}{\partial s_j^{(k)}} \quad (\text{B.13})$$

e como as derivadas parciais para $i \neq j$ se anulam

$$\delta_j^{(k)} = -\frac{1}{2} \frac{\partial (d_j - f(s_j^{(k)}))^2}{\partial s_j^{(k)}} = -(d_j - f(s_j^{(k)})) \frac{\partial (d_j - f(s_j^{(k)}))}{\partial s_j^{(k)}} = (d_j - x_j^{(k)}) f'(s_j^{(k)}) \quad (\text{B.14})$$

Erro na saída associado ao neurônio j da última camada:

$$\varepsilon_j^{(k)} = d_j - x_j^{(k)} = d_j - y_j \quad (\text{B.15})$$

implicando em

$$\delta_j^{(k)} = \varepsilon_j^{(k)} f'(s_j^{(k)}) \quad (\text{B.16})$$

Desenvolvimento para uma camada oculta (k):

O erro quadrático da camada k é determinado pelas saídas lineares da camada $k+1$.

$$\delta_j^{(k)} = -\frac{1}{2} \frac{\partial \varepsilon^2}{\partial s_j^{(k)}} \quad (\text{B.17})$$

O erro quadrático da camada k é determinado pelas saídas lineares da camada $k+1$.

$$\begin{aligned} \delta_j^{(k)} &= \frac{1}{2} \frac{\partial \varepsilon^2}{\partial s_j^{(k)}} = -\frac{1}{2} \sum_{i=1}^{N_{k+1}} \left(\frac{\partial \varepsilon^2}{\partial s_i^{(k+1)}} \frac{\partial s_j^{(k+1)}}{\partial s_i^{(k)}} \right) \quad (\text{regra da cadeia}) \\ &= \sum_{i=1}^{N_{k+1}} \left(\left(-\frac{1}{2} \frac{\partial \varepsilon^2}{\partial s_i^{(k+1)}} \right) \frac{\partial s_i^{(k+1)}}{\partial s_i^{(k)}} \right) = \sum_{i=1}^{N_{k+1}} \left(\delta_i^{(k+1)} \frac{\partial s_i^{(k+1)}}{\partial s_j^{(k)}} \right) \end{aligned} \quad (\text{B.18})$$

lembrando que $s_j^{(k)} = w_{0j}^{(k)} + \sum_{i=1}^{N_i} w_{ij}^{(k)} x_i^{(k-1)}$, tem-se

$$\delta_j^{(k)} = \sum_{i=1}^{N_{k+1}} \left(\delta_i^{(k+1)} \frac{\partial}{\partial s_i^{(k)}} \left(w_{oi}^{(k+1)} + \sum_{l=1}^{N_k} w_{li}^{(k+1)} f(s_l^{(k)}) \right) \right) \quad (\text{B.19})$$

observando que $\frac{\partial}{\partial s_j^{(k)}} f(s_l^{(k)}) = 0$ se $l \neq j$ e que $\frac{\partial}{\partial s_j^{(k)}} f(s_j^{(k)}) = f'(s_j^{(k)})$ temos:

$$\delta_j^{(k)} = \sum_{i=1}^{N_{k+1}} (\delta_i^{(k+1)} w_{ji}^{(k+1)} f'(s_j^{(k)})) \quad (\text{B.20})$$

Definindo-se $\varepsilon_j^{(k)} = \sum_{i=1}^{N_{k+1}} \delta_i^{(k+1)} w_{ji}^{(k+1)}$, tem-se o erro derivativo quadrático como:

$$\delta_j^{(k)} = \varepsilon_j^{(k)} f'(s_j^{(k)}) \quad (\text{B.21})$$

A atualização dos pesos é feita por

$$W_j^{(k)}(n+1) = W_j^{(k)}(n) + \mu(-\vec{\nabla} W_j^{(k)}(n)) \quad (\text{B.22})$$

$\mu > 0 \rightarrow$ taxa de aprendizagem

$n \rightarrow$ iteração corrente

Temos então:

$$W_j^{(k)}(n+1) = W_j^{(k)}(n) + 2\mu\delta_j^{(k)}(n)X_j^{(k)}(n) \quad (\text{B.23})$$

O algoritmo *backpropagation*

1. $w_{ij}^{(k)} \leftarrow \text{random}$, inicializar a rede
2. para (x,d) , par de treinamentos, obter y . Propagação feedforward. $\varepsilon^2 = \sum_{j=1}^m (d_j - y_j)^2$
3. $k \leftarrow$ última camada
4. para todo elemento j da camada k faça
 Calcule $\varepsilon_j^{(k)}$ empregando $\varepsilon_j^{(k)} = d_j - x_j^{(k)} = d_j - y_j$ se k for a última camada,

$$\varepsilon_j^{(k)} \sum_{i=1}^{N_{k+1}} \delta_i^{(k+1)} w_{ji}^{(k+1)} \quad \text{se for uma camada oculta;}$$

$$\text{Calcule } \delta_j^{(k)} = \varepsilon_j^{(k)} f'(s_j^{(k)})$$

5. $k \leftarrow k-1$ se $k > 0$ vá para o passo 4, senão prossiga.
6. $W_j^{(k)}(n+1) = W_j^{(k)}(n) + 2\mu\delta_j^{(k)}(n)X_j^{(k)}(n)$
7. para o próximo par de treinamento vá para o passo 2.

O algoritmo *backpropagation* levou a uma grande aceitação das RNAs por parte da comunidade científica, uma vez que redes multicamadas podem ser treinadas a partir dos dados que representam amostras significativas do processo em questão.

Nesta forma mostrada anteriormente algoritmo *backpropagation* é em geral muito lento e suscetível a “patologias” de treinamento. A *paralisia da rede* ocorre em regiões de gradiente próximos de zero (platôs). Dependendo das condições iniciais o treinamento fica preso em *mínimos locais* da superfície de erro.

O critério de parada do treinamento considera em geral um limite máximo do número de épocas de treinamento. Além disso, o treinamento pode terminar quando a soma do erro quadrático ou a média deste atinge o seu objetivo.

Para acelerar a convergência utilizam-se variantes do *backpropagation*. A função *trainbpm* (com momento, m_c) considera não apenas o gradiente local, mas também tendências recentes da superfície de erro. Atua como um filtro passa-baixas, ignorando características

menores da superfície de erro. O treinamento “desliza” sobre mínimos locais não muito pronunciados.

$$\Delta W_j^{(k)}(n+1) = m_c \Delta W_j^{(k)}(n) + (1 - m_c) 2\mu \delta_j^{(k)}(n) X_j^{(k)}(n) \quad (\text{B.24})$$

$$m_c \cong 0,95 \text{ (típico)}$$

A função *trainbpx* implementa uma taxa de aprendizagem adaptativa (e com momento, m_c).

Taxa de aprendizagem: $1,04 \frac{\text{erro novo}}{\text{erro anterior}}$ redução de μ : 0,7 aumento de μ : 1,05

O algoritmo de *Levenberg-Marquardt* é um dos mais rápidos; utiliza a Matriz J Jacobiana das derivadas do erro (e) em relação aos pesos.

$$\Delta W_j^{(k)} = (J^T J + \mu J)^{-1} J^T \quad (\text{B.25})$$

Se o fator de escala μ é muito grande temos o algoritmo do gradiente descendente. Se μ é muito pequeno torna-se o algoritmo de Gauss-Newton, que é mais preciso e assim adequado para uma rápida convergência próxima a um mínimo.

B.2 Rede Neural Probabilística (PNN – *Probabilistic Neural Network*)

Considerações iniciais: uma imagem formada com texturas T_Ω designada por um domínio $\Omega = \{(i, j), 0 \leq i < I, 0 \leq j < J\}$ de posições de *pixels*. Considerando $\{g_s \in R^M, \forall s \in \Omega\}$ um conjunto de vetores de características M -dimensional utilizados para definir a imagem, onde cada g_s é um vetor M -dimensional que caracteriza o *pixel* s em Ω . O g_s pode ser considerado como uma instância do processo aleatório G_s , L_s é a variável aleatória descrevendo o rótulo do *pixel* s . Assumindo que L_s pode ter qualquer valor do conjunto $\{0, 1, \dots, K-1\}$ onde K é o número de classes de texturas. A classe correspondente são denominadas por $C_0 \dots C_{k-1}$. Sendo Ω_k , um subconjunto de Ω , o conjunto de treino de C_k . A notação $\text{card}(\Omega)$ é usada para a cardinalidade de qualquer conjunto Ω . Fazendo N_s^p ser um conjunto de *pixels* na p -ésima ordem dos vizinhos simétricos do *pixel* s .

O processo que formula a probabilidade de um conjunto de vetores de um *pixel* s dado pelo modelo de parâmetros de cada classe k é dada por

$$P(G_s = g_s | L_s = k) = \frac{e^{-E_f(G_s = g_s | L_s = k)}}{Z_f} \quad (\text{B.26})$$

onde E_f é uma função energia definida pelo modelo estatístico selecionado e Z_f é uma constante de normalização considerada independente de s e k .

O rótulo de cada *pixel* na imagem dependem dos rótulos dos pixels de seus vizinhos. O processo $P(L_s | L_r, \forall r \in N_s^p)$ descreve a probabilidade de um rótulo de cada *pixel* s dados os rótulos dos *pixels* na p -ésima ordem uniforme dos vizinhos de N_s^p de s

$$P(L_s, | L_r, \forall r \in N_s^p) = \frac{\exp \left[\sum_{\forall r \in N_s^p} \beta \delta(L_s - L_r) \right]}{Z_p} \quad (\text{B.27})$$

onde β é uma constante positiva, $\delta(\cdot)$ é a função delta de Kronecker²⁰ e Z_p é uma constante de normalização.

O processo de competição tenta reduzir a probabilidade de um rótulo ser assumir o valor de outro e é expresso como

$$P(L_s = k | L_s) = \frac{\exp \left[-\alpha \sum_{l \in \tilde{L}_s} \bar{\delta}(k - l) \right]}{Z_c} \quad (\text{B.28})$$

onde \tilde{L}_s é o conjunto de rótulos que podem ser designados ao *pixel* s e Z_c é uma constante de normalização.

A rede PNN é uma rede com três camadas consistindo em uma camada de entrada, uma camada interna e uma camada de soma, possuindo a mesma estrutura da rede radial mostrada na Figura B.3. A camada de entrada contém M neurônios de forma aceitar um vetor de características M -dimensional. A camada interna consiste em K neurônios, sendo que o k -ésimo neurônio da camada de entrada contém S_k números de neurônios, onde $S_k = \text{card}(\Omega_k)$. Cada neurônio na camada interna é conectado a todos os neurônios da camada de entrada. A camada de soma consiste em K neurônios, um para cada neurônio da camada interna. Uma

²⁰ Notação $\delta(i,j)$ definida por: $\delta(i,j) = 1$ se $i = j$; $\delta(i,j) = 0$ se $i \neq j$.

função radial e uma função de ativação Gaussiana são usadas nos neurônios da camada intermediária. Para a camada de saída (somatório), é utilizada uma função de ativação .

Na rede cada vetor de treinamento g_m ($0 \leq m < S_k$) da classe C_k é armazenado com o valor $W_{m,k}$ conectando os neurônios da camada de entrada ao m-ésimo neurônio da camada intermediária. O peso da conexão de cada neurônio da camada intermediária e os da camada de soma para a k-ésima classe é definido como $1/S_k$.

Para cada vetor de entrada $g_s, s \in \Omega$, a saída do m-ésimo neurônio é $e^{-\|g_s - W_{m,k}\|^2 / \sqrt{(2\pi)^M \sigma_{m,k}^2}}$, onde $\sigma_{m,k}$ é um parâmetro para a função Gaussiana de ativação daquele neurônio. O processo de formação dado por (B.26) é descrito como

$$P(G_s = g_s | L_s = k) = \frac{1}{S_k} \sum_{m=1}^{S_k} \frac{e^{-\|g_s - W_{m,k}\|^2 / 2\sigma_{m,k}^2}}{\sqrt{(2\pi)^M \sigma_{m,k}^2}} = \exp \left\{ \ln \left[\frac{1}{S_k} \sum_{m=1}^{S_k} \frac{e^{-\|g_s - W_{m,k}\|^2 / 2\sigma_{m,k}^2}}{\sqrt{(2\pi)^M \sigma_{m,k}^2}} \right] \right\}. \quad (B.29)$$

Essa relação mostra que a rede PNN representa o espaço de vetores de entrada de cada classe C_k como uma distribuição de Modelo de Mistura Gaussiana²¹ (*Gaussian Mixture Model - GMM*), onde existem tantos componentes quantos forem conjuntos de treino para aquela classe. Cada componente m da mistura da classe C_k é representado pelo m-ésimo vetor de treinamento g_m como média e a constante $\sigma_{m,k}^2$ como variância.

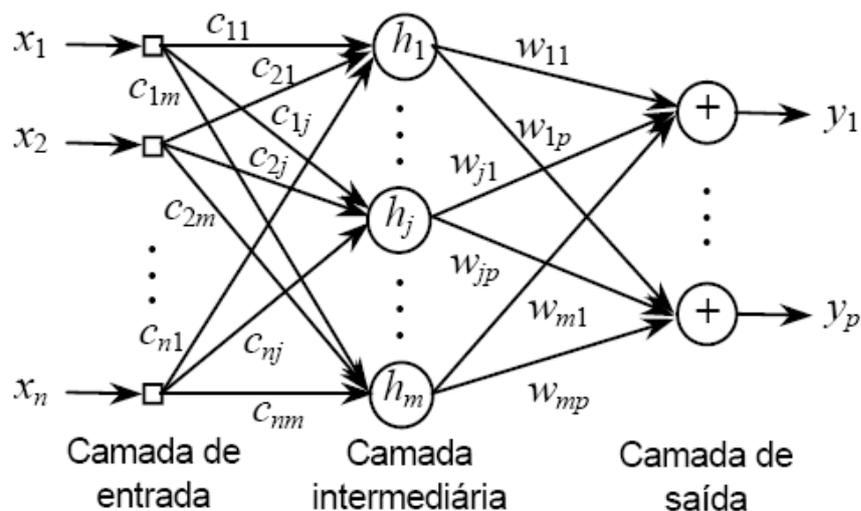


Figura B.3 – Arquitetura básica de uma PNN

²¹ Modelo cujas variáveis independentes são frações de um total.

Apêndice C

Conectividade e Medidas de Irregularidades

C.1 Conectividade

Conectividade em imagens digitais introduz um conjunto de teorias e algumas topologias, demonstrando o uso de termos como caminhos (*paths*), *4-conectado*, etc. Definindo S um subconjunto de pontos inteiros (i, j) de uma imagem, um *4-caminho* é qualquer N agrupamento de pontos (i_n, j_n) , $n = 1, \dots, N$ onde $N > 1$, e para qualquer par n (i_n, j_n) , (i_{n+1}, j_{n+1}) são vizinhos horizontais ou verticais. Da forma

$$|(i_n - i_{n+1}) + (j_n - j_{n+1})| \leq 1, \forall n. \quad (C.1)$$

Similarmente, um 8-caminho é qualquer agrupamento de N pontos (i_n, j_n) , $n = 1, \dots, N$ onde $N > 1$, e para qualquer par n (i_n, j_n) , (i_{n+1}, j_{n+1}) são vizinhos horizontais, verticais ou diagonais. Da forma

$$\max\{|i_n - i_{n+1}|, |j_n - j_{n+1}|\} \leq 1, \forall n. \quad (C.2)$$

Qualquer par de pontos $(i_1, j_1), (i_2, j_2)$ em S são ditos *4-conectados* se existe em S um *4-caminho* tendo (i_1, j_1) como primeiro ponto e (i_2, j_2) como ultimo ponto. Se existir em S um *4-caminho* contendo todos os elementos de S , então S é chamado de *4-conectado*, se possuir somente um componente *4-conectado* recebe o nome de *4-componente*, o termo *8-conectado* pode ser definido da mesma maneira.

Na terminologia de SGF, um *4-componente* S em uma imagem binária I_b é simplesmente uma região cuja intensidade do *pixel* são todas de valor '1' ou de valor '0', Da forma

$$I_b(i, j) = 1, \forall (i, j) \in S, \quad (C.3)$$

ou

$$I_b(i, j) = 0, \forall (i, j) \in S. \quad (C.4)$$

Um *pixel* p é conectado a um *pixel* q se p é *4-conectado* a q ou se p é *4-conectado* a um terceiro *pixel* que também é conectado a q . Ou, em outras palavras, dois *pixels* q e p são conectados se existir um caminho de p a q no qual cada *pixel* também é *4-conectado*.

Um exemplo de imagem binária com dois componentes *4-conectados* pode ser vista na Figura C.1. Se a conectividade for baseada em 8 vizinhos (*8-conectado*), os dois componentes podem ser mesclados em um.

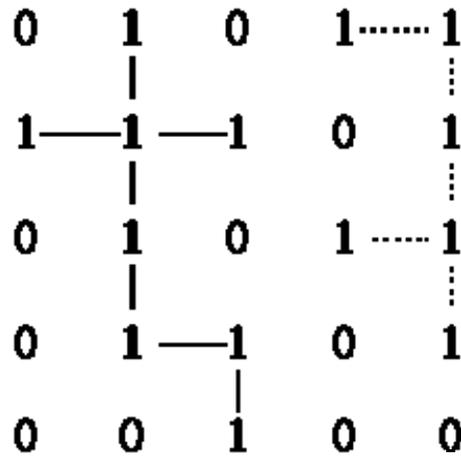


Figura C.1 – Dois componentes 4-conectados.

C.2 Medidas de Irregularidades

Muitas características das medidas de IRGL merecem algumas considerações. Uma medida de uma região estritamente relacionada com irregularidade é a “circularidade” ou “compacidade”, baseada no perímetro P a na área A de uma região,

$$circularidade = \frac{P}{\sqrt{4\pi A}}. \quad (C.5)$$

Pode ser visto que para uma região circular de raio r ,

$$circularidade = \frac{2\pi r}{\sqrt{4\pi}\sqrt{A}} = \frac{2\pi r}{\sqrt{4\pi}\sqrt{\pi r^2}} = 1. \quad (C.6)$$

O resultado pouco desejado dessa definição demonstra que quanto maior a medida de circularidade menor será a indicação da mesma. Redefinindo e renomeando esta medida para “irregularidade”, tem-se uma forma mais apropriada. Da equação (C.6), a irregularidade de uma região circular é agora definida por:

$$irregularidade = \frac{2\pi r}{\sqrt{4\pi}\sqrt{A}} - 1 = \frac{\sqrt{\pi}r}{\sqrt{A}} - 1, \quad (C.7)$$

onde r é definido como a distância do centro de gravidade da região ao ponto mais afastado do perímetro,

$$r = \sup_{(x,y) \in A} \sqrt{(x - \bar{x})^2 + (y - \bar{y})^2}, \quad (C.8)$$

$$\bar{x} = \int_A x dx, \quad \bar{y} = \int_A y dy. \quad (C.9)$$

O termo -1 define a irregularidade de uma área circular a ter valor 0.

Em uma grade digital, as equações (C.7) e (C.8) se tornam:

$$irregularidade = \frac{\sqrt{\pi} \max_{i \in R} \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}}{\sqrt{A}} - 1 \quad (C.10)$$

onde R é o conjunto de índices dos *pixels* na região. A área da região A pode ser definida como

$$A = |R|, \quad (C.11)$$

a cardinalidade de R (ou o número de *pixels* na região).

Usando uma grade digital podem ocorrer problemas quando de regiões contendo somente um único *pixel*:

$$irregularidade = \frac{\sqrt{\pi} \max_{i \in R} \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}}{\sqrt{|R|}} - 1 = \frac{\sqrt{\pi} \cdot 0}{1} - 1 = -1. \quad (C.12)$$

Chen et al. (1984) superaram este problema pela adição de um termo ‘1’ ao numerador,

$$IRGL = \frac{1 + \sqrt{\pi} \max_{i \in R} \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}}{\sqrt{|R|}} - 1, \quad (C.13)$$

onde

$$\bar{x} = \frac{\sum_{i \in R} x_i}{|R|}, \quad \bar{y} = \frac{\sum_{i \in R} y_i}{|R|}. \quad (C.14)$$

Definida dessa maneira a medida de irregularidade possui um número distinto de propriedades:

1. Para uma região de um único *pixel*, a medida assume a forma,

$$IRGL = \frac{1 + \sqrt{\pi} \cdot 0}{1} - 1 = 0. \quad (C.15)$$

2. Conforme o raio de uma região circular tende a infinito, ou alternativamente, conforme o espaçamento da grade ε se aproxima de 0,

$$\begin{aligned} IRGL &= \lim_{\varepsilon \rightarrow 0} \left(\frac{1 + \sqrt{\pi} \max_{i \in r} \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}}{\sqrt{|R|}} - 1 \right) \\ &= \lim_{\varepsilon \rightarrow 0} \left(\frac{1}{\sqrt{|R|}} + \sqrt{\pi} \frac{\max_{i \in r} \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}}{\sqrt{|R|}} - 1 \right) \\ &= 0 + 1 - 1 = 0. \end{aligned} \quad (C.16)$$

3. A medida IRGL é invariante sobre rotação e translação.

Isto pode ser facilmente visto porque $|R|$ e $\max_{i \in r} \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}$ permanecem constantes.

Apêndice D

Biblioteca VisTex

D.1 Biblioteca VisTex

A biblioteca VisTex é uma coleção de imagens de texturas. Criada com a intenção de prover um conjunto de texturas coloridas de alta qualidade para aplicações de visão computacional.

As imagens de texturas utilizadas são mostradas na Figura D.1, extraídas do endereço: <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html>.



Figura D.1 – Texturas utilizadas, extraídas da biblioteca VisTex.